

UNIVERSITÉ DES ANTILLES ET DE LA GUYANE

## MÉMOIRE DE SYNTHÈSE

pour l'obtention d'une

## HABILITATION À DIRIGER DES RECHERCHES

par

Vivien ROSSI

Modélisation des forêts tropicales en intégrant leur  
diversité spécifique et structurelle

présenté le 22 mai 2013 devant un jury composé de :

M. Baraloto Christopher, Directeur de recherches, Inra-UMR Ecofog, Guyane  
M. Bar-Hen Avner, Professeur, Université de Paris-Descartes, Paris 5  
M. Courbaud Benoît, Chargé de recherches HDR, Irstea-Cemagref, Saint-Martin d'Hères  
M. Hérault Bruno, Chargé de recherches, Cirad-UMR Ecofog, Guyane  
M. Pelissier Raphaël, Directeur de recherches, IRD-UMR Amap, Montpellier



# Table des matières

<b>I</b>	<b>Présentation du candidat</b>	<b>5</b>
<b>1</b>	<b>Curriculum vitae</b>	<b>7</b>
1.1	Titres universitaires . . . . .	7
1.2	Fonctions occupées . . . . .	8
1.3	Responsabilités . . . . .	9
1.4	Enseignements . . . . .	10
1.5	Projets de Recherche . . . . .	12
<b>2</b>	<b>Liste des publications</b>	<b>17</b>
2.1	Articles en préparation . . . . .	17
2.2	Articles publiés dans des revues avec facteur d'impact . . . . .	17
2.3	Articles publiés dans des revues à comité de lecture, sans facteur d'impact . . . . .	20
2.4	Chapitre de livre . . . . .	20
2.5	Actes de conférences . . . . .	20
2.6	Rapports de recherche . . . . .	21
2.7	Communications dans des congrès internationaux . . . . .	21
2.8	Communications dans des congrès nationaux . . . . .	24
<b>3</b>	<b>Encadrement d'étudiants</b>	<b>26</b>
3.1	Thèses . . . . .	26
3.2	Post-Doctorats . . . . .	28
3.3	Stages de DEA et Master 2 . . . . .	30
3.4	Stages d'élèves ingénieurs . . . . .	31
3.5	Stages de Master 1 . . . . .	31
3.6	Comités de thèse . . . . .	32

<b>II</b>	<b>Travaux de recherche</b>	<b>33</b>
<b>1</b>	<b>Introduction</b>	<b>35</b>
<b>2</b>	<b>Méthodologie de modélisation statistique</b>	<b>37</b>
2.1	Inférence de modèles de systèmes dynamiques . . . . .	37
2.1.1	Modèles de Markov cachés . . . . .	38
2.2	Inférence de modèles statistiques . . . . .	42
2.2.1	Algorithme MCMC en parallèle et en interaction . . . . .	43
2.3	Sélection de modèles multi-réponses . . . . .	45
2.4	Conclusion . . . . .	46
<b>3</b>	<b>Comment modéliser les forêts tropicales en intégrant leur diversité spécifique ?</b>	<b>48</b>
3.1	Évaluation de la biomasse forestière . . . . .	49
3.1.1	Estimation de la biomasse d'une parcelle à partir d'inventaires d'exploitants forestiers . . . . .	51
3.1.2	Estimation de la biomasse d'une parcelle à partir d'inventaires scientifiques . . . . .	53
3.2	Modélisation de la dynamique des forêts tropicales à l'échelle de la communauté . . . . .	58
3.2.1	Approche par groupes fonctionnels . . . . .	60
3.2.2	Approche par les traits fonctionnels . . . . .	66
3.2.3	Groupes fonctionnels versus Traits fonctionnels . . . . .	74
3.3	Vers une quantification de l'impact des changements climatiques sur les forêts tropicales . . . . .	77
<b>4</b>	<b>Perspectives</b>	<b>86</b>
4.1	Méthodologie en modélisation statistique . . . . .	86
4.2	Modèles de dynamiques forestières climat dépendant . . . . .	88
4.3	Prise en compte des enjeux socio-économiques des forêts tropicales . . . . .	90

**Première partie**

**Présentation du candidat**



# 1 Curriculum vitae

Vivien Rossi  
Né le 12 juin 1976 à Montpellier  
Marié, 3 enfants

Cirad - UMR Ecofog  
Campus Agronomique, BP 701  
97387 Kourou Cedex  
vivien.rossi@cirad.fr

## 1.1 Titres universitaires

- 2005 Qualification aux fonctions de Maître de conférences en section 26
  
- 2004 Doctorat en Biostatistique, Montpellier-SupAgro, Mention Très Honorable  
Intitulé : Filtrage non linéaire par noyaux de convolution. Application à un procédé de dépollution biologique.  
Sous la direction de Jean-Pierre Vila, DR INRA dans l'UMR Analyse des Systèmes et Biométrie.
  
- 2001 DEA de Biostatistique, Université de Montpellier 2, Mention Bien
  
- 2000 Maîtrise de Mathématiques MIM, Université de Montpellier 2, Mention Bien

## 1.2 Fonctions occupées

- Depuis 08/2009    Cirad Guyane, chercheur dans l'UMR Ecofog  
Modélisateur de la dynamique des forêts tropicales humides
- 01/2007 - 08/2009    Cirad Montpellier, chercheur dans l'UR Dynamiques des forêts naturelles  
Modélisateur de la dynamique des forêts tropicales humides
- 09/2006 - 01/2007    Université de Perpignan, Attaché Temporaire d'Enseignement et de Recherche à l'IUT STID  
Chercheur associé à l'Institut Universitaire de Recherche Clinique de l'Université Montpellier 1
- 09/2005 - 08/2006    Inria - Irisa Rennes, Post-doctorant dans le projet de recherche Aspi  
Recherche sur la modélisation bayésienne et markovienne pour le vivant. Application à l'évaluation et la gestion de ressources naturelles
- 09/2004 - 08/2005    Université Montpellier 2, Attaché Temporaire d'Enseignement et de Recherche dans l'UMR Institut de Mathématiques et de Modélisation de Montpellier  
Recherche sur l'estimation de paramètres inconnus dans les systèmes dynamiques partiellement observés. Application aux modèles de dynamique des populations de vertébrés
- 11/2001 - 12/2004    Inra France, Doctorant dans l'UMR ASB, Montpellier  
Recherche sur les systèmes dynamiques non linéaires partiellement observés. Application à un bioprocédé de dépollution d'eaux usées



## 1.3 Responsabilités

Depuis 05/2011	Responsable du laboratoire de modélisation de l'UMR Ecofog : 5 chercheurs, 2 doctorants, 3 ingénieurs, 5 techniciens et 1 ouvrier
Depuis 10/2011	Correspondant Cirad pour l'UMR Ecofog : interlocuteur de la direction scientifique, gestion du personnel et du budget
06/2011 - 01/2013	Gestionnaire du réseau de parcelles Guyafor : plus de 235 hectares de forêts, sur 12 dispositifs, inventoriés régulièrement.

## 1.4 Enseignements

- 2010-2012 Cours Outils statistiques pour l'écologie, Master 2 Ecologie des Forêts Tropicales, Université Antilles-Guyane et AgroParisTech (16h/an)
- 2010-2011 Cours Méthodes d'ordination pour l'analyse des données écologiques, Ecole thématique CNRS - Méthodes et modèles pour l'étude de la biodiversité amazonienne, Kourou (3h)  
Cours d'introduction aux méthodes de Monte Carlo, UMR Ecofog, Kourou (6h)
- 2006 - 2009 Cours d'introduction aux méthodes de Monte Carlo, Master 2 de Biostatistique, Université de Montpellier 2 (3h/an)
- 2007 - 2009 Cours de régression linéaire pour les dynamiques des forêts tropicales, 2ème année ingénieur agronome, Montpellier-SupAgro (3h/an)
- 2006-2007 Atelier thématique : Probabilités, statistique, Scilab - Quelques applications en environnement Université de Fianarantsoa - Madagascar (40 h)  
Cours Régression Linéaire et Dynamique des populations, 2nd année, Montpellier-SupAgro (3h)

- 2006-2007 Cours Statistique Descriptive, 1ère année IUT STID, Université de Perpignan (16h)  
 TP Découverte des logiciels de statistique, 1ère année IUT STID Université de Perpignan (45h)  
 TD Estimation et Tests, 2nd année IUT STID, Université de Perpignan (45h)
- 2005-2006 TD Régression linéaire, 2nd année ingénieur, ENSAI, Rennes (60h)
- 2004-2005 TD et TP de Biostatistique, 2nd année IUP BBB, Université de Montpellier 2 (18h+27h)  
 Cours et TD de Modèles de Survie, Licence MASS, Université de Montpellier 2 (18h+18h)  
 TD de Statistique Mathématique, Licence MASS, Université de Montpellier 2 (18h)  
 TD et TP d'Initiation à la Statistique, L1, Université de Montpellier 2 (15h+30h)  
 Cours Modèles de Markov cachés et Filtre de Kalman, Préparation à l'Agrégation de Mathématiques, Université de Montpellier 2 (6h)
- 2003-2005 Formation au traitement statistique d'enquêtes, Mastère DAT, Cnearc (24h/an)  
 Encadrement du traitement statistique d'une enquête, Mastère DAT, Cnearc (24h/an)
- 2002-2004 TD de Statistique, licence de psychologie, Université de Montpellier 3 (44 h/an)
- 2001-2003 TD Mathématiques financières et Statistique, 1ère année Deug AES, Univ. Montpellier 3 (23h/an)  
 TD Méthodologie mathématique, 1ère année Deug AES, Université de Montpellier 3 (20h/an)

## 1.5 Projets de Recherche

### **GuyaSim : un simulateur pour explorer l'impact des scénarios de développement de la Guyane sur les services des écosystèmes forestiers**

Niveau d'implication : responsable du projet

Partenaires : UMR Ecofog, UR B&Sef, UMR Lerfob, EPI Mere Inria

Subvention : 500k€, PO Feder Guyane, 2011-2013

Lien : <http://www.ecofog.gf/spip.php?article429>

L'objectif principal de ce projet est construire un « simulateur de la Guyane » permettant de quantifier et de cartographier les évolutions probables d'indicateurs de services environnementaux, tels que le stock de carbone, la beta-biodiversité ou le fonctionnement du sol selon différents scénarios prédéfinis. Nous le faisons en mutualisant et en complétant les résultats des travaux scientifiques obtenus depuis plusieurs années par différents organismes de recherche sur la forêt guyanaise. Cet outil aura une double fonctionnalité, d'une part il constituera un outil d'aide à la décision permettant d'optimiser l'aménagement et le développement du territoire, d'autre part il fournira un laboratoire virtuel aux scientifiques permettant de tester plus facilement certaines hypothèses écologiques.

Le plan d'action se décompose en 4 activités :

**Activité 1** Spatialisation des services écosystémiques à l'échelle de la forêt guyanaise : détermination des indicateurs des services écosystémiques et leur évaluation à l'échelle de la Guyane selon les différents types forestiers ou agricoles

**Activité 2** Scénarios socio-économiques et dynamiques territoriales : concerne les questions relatives aux aspects socio-économiques du développement guyanais

**Activité 3** Scénarios d'impacts des changements climatiques pour les forêts guyanaises : modélisation de l'impact des scénarios de changements climatiques

**Activité 4** Réalisation du logiciel de simulation : développement d'un logiciel, avec une interface graphique pour les utilisateurs, intégrant les résultats et connaissances des trois autres activités

## **ClimFor : Modélisation des impacts des changements climatiques sur la biodiversité de la forêt tropicale de Guyane Française**

Niveau d'implication : responsable de l'activité 3

Responsable du projet : Bruno Hérault, Cirad-UMR Ecofog Partenaires : UMR EcoFoG, ONF Guyane, UMR EDB, UMR EEF, DSGE-Université de Liège.

Subvention : 250k€, programme « Modélisation et scénarios de biodiversité » de la FRB, 2011-2013

Lien : <http://www.ecofog.gf/spip.php ?article113>

Ce projet vise à explorer les conséquences du changement climatique sur quatre services écosystémiques (diversité végétale, diversité fonctionnelle, stockage de carbone, ressource en bois). En intégrant des données de différentes sources à travers des approches de modélisation originales, nous serons en mesure de modéliser les relations entre 3 composantes du changement climatique (stress hydrique, rayonnement incident, température) et la valeur des services écosystémiques. Nous pourrions ainsi explorer les conséquences d'un ensemble de scénarios de changements climatiques sur la valeur de ces services écosystémiques. Nous examinerons aussi l'adéquation entre le réseau des aires protégées dans le nord de la Guyane et la résilience de leurs services écosystémiques.

Le plan d'action se décompose en 5 activités :

**Activité 1** Quel est le niveau des variations spatio-temporelles des précipitations (stress hydrique), de l'irradiance et de la température ?

**Activité 2** Comment varient les services écosystémiques de la forêt guyanaise selon les paysages ?

**Activité 3** Dans quelle mesure la variation spatiale des indicateurs climatiques affecte les services écosystémiques ? Avec quelle précision la relation entre le climat et les indicateurs écosystémiques peut être utilisée pour interpoler une carte fondée sur les habitats ?

**Activité 4** Est-ce que les indicateurs climatiques dynamiques affectent la dynamique forestière au travers des processus de croissance et de mortalité des arbres ? Et dans quelle mesure les valeurs prédites des services écosystémiques par les modèles de dynamique forestière intégrant le climat, vont-elles varier ?

**Activité 5** Peut-on optimiser le réseau des aires protégées en Guyane française pour tenir compte de tous les avenir possibles ? Quelles règles d'exploita-

tion doivent être mises en place par les gestionnaires forestiers pour maintenir à long terme des populations d'arbres viables ?

### **GuyaSpaSE : Comment passer des estimations locales de biodiversité et de stocks de carbone à des indicateurs régionaux utilisables dans l'aménagement et la gestion des massifs forestiers guyanais ?**

Niveau d'implication : co-responsable de l'activité 2

Responsable du projet : Christopher Baraloto, Inra-UMR Ecofog

Partenaires : UMR Ecofog, ONF, Herbier de Cayenne, UMR Amap

Subvention : 100k€, programme BGF du Gip Ecofor, 2011-2013

Lien : <http://www.ecofog.gf/spip.php?article547>

Dans ce projet nous proposons de définir des outils fiables d'évaluation de la biodiversité à l'échelle de travail des gestionnaires. Pour ce faire, nous proposons d'aborder la biodiversité à travers la notion d'« habitat » qui est la plus appropriée, eu égard à son échelle de perception, à son caractère intégrateur et à son utilisation partagée par les différentes disciplines (botanistes, géographes, forestiers, spécialistes de la faune. . .). Trois axes de travail seront privilégiés : trouver les indices les plus pertinents pour capter les différents patrons de diversité et de carbone, définir des protocoles de collecte de données sur le terrain, rapides et répondant conjointement aux deux services et étudier la variabilité spatiale des habitats forestiers à l'échelle des paysages en Guyane.

Le plan d'action se décompose en 3 activités :

**Activité 1** Estimateurs des services environnementaux pour l'aménagement forestier

**Activité 2** Méthodes pour des inventaires rapides en forêt tropicale humide

**Activité 3** Variabilité spatiale des services environnementaux des forêts guyanaises

## **Nouragues : Examining tropical tree community structure and dynamics at appropriate spatial scales**

Niveau d'implication : Analyses de données et traitements statistiques

Responsable du projet : Christopher Baraloto, Inra-UMR Ecofog

Partenaire : UMR Ecofog

Subvention : 15 k€, programme CNRS - Nouragues, 2011

Ce projet visait à tester des hypothèses écologiques (limitation de la dispersion, le filtrage de l'environnement, . . . ) en s'appuyant sur les données d'une placette de 10ha sur le site des Nouragues. L'originalité de l'étude résidait dans l'utilisation des traits fonctionnels pour appréhender la diversité spécifique dans les modèles.

## **Mirc : Modélisation stochastique, inférence numérique et contrôle pour l'évaluation et la gestion de ressources renouvelables**

Niveau d'implication : correspondant du partenaire Cirad

Responsable du projet : Fabien Campillo, Inria-Projet Mere

Partenaires : Projet Mere et Projet Virtual Plants Inria, Cermics ENPC, UMR ASB, Cirad UR37 et UR39

Subvention : 80k€, Actions de Recherche Coopérative de l'Inria, 2007-2008

Ce projet visait à développer des modèles stochastiques pour la gestion de ressources renouvelables. Les applications du projet concernaient les forêts tropicales naturelles et les pêcheries. Pour les forêts tropicales, nous avons développé un modèle individu centré (IBM), en temps continu.

## **Mip : Modélisation Intégrée en Dynamique des Populations : Applications à la Gestion et à la Conservation**

Niveau d'implication : Responsable scientifique du partenaire UM2

Responsable du projet : Roger Pradel, CNRS-UMR Cefe

Partenaires : CNRS-UMR Cefe, Inra-UMR ASB et UM2-UMR I3M

Subvention : 25k€, Action Incitative Régionale (2005-2006)

L'objectif du projet était de développer des approches et outils généraux de modélisation intégrée de la dynamique des populations de vertébrés, à des fins de

gestion et de conservation. Nous avons développé des approches et algorithmes originaux pour calibrer les modèles de capture-recapture. Ces méthodes ont été ajoutées au logiciel M-SURGE, développé au CEFÉ, permettant le transfert vers les biologistes des populations.

### **AS CNRS : Observation des entrées inconnues pour l'aide à la conduite et au diagnostic de procédés biologiques de traitement des eaux usées et déchets**

Niveau d'implication : participant

Responsable du projet : Isabelle Quiennec, CNRS-UMR Laas

Partenaires : Projet Comore-Inria, LBE-Inra, Lipe-Insa, Inra-UMR ASB et Laas-CNRS

Subvention : 20 k€, Action Spécifique CNRS, 2003-2004

Ce projet se focalisait sur le développement et l'utilisation de capteurs logiciels pour aider au diagnostic des installations de type bioréacteur. Il s'agissait de reconstruire l'état du système lorsque l'entrée n'est pas connue. De plus, les paramètres variant dans le temps, l'absence de mesures fiables et les fortes variations des conditions opératoires et des charges de pollution imposaient le développement et l'utilisation de techniques avancées de filtrage non linéaire.



## 2 Liste des publications

### 2.1 Articles en préparation

1. M. Aubry-Kientz, B. Hérault, C. Ayotte-Trepanier, C. Baraloto, V. Rossi, *Towards trait-based mortality models for tropical forests ?*, en révision pour PLoS ONE
2. F. Wagner, V. Rossi, C. Baraloto, D. Bonal, C. Stahl, B. Hérault, *Functional traits explain tropical tree growth response to climate*, soumis à Functional Ecology
3. Q. Molto, B. Hérault, O. Brunaux, G. Coste, V. Freycon, S. Guitet, L. Blanc, V. Rossi, *Mapping Biomass of tropical forests : uncertainty propagation with a model-base approach*, en préparation
4. Q. Molto, B. Hérault, JJ Boreux, M. Daullet, A. Rousteau, V. Rossi, *Predicting tree height for biomass estimates in tropical forest*, en préparation
5. A. Dejean, J. Orivel, V. Rossi, O. Roux, P.-J. Malé, R. Céréghino, Céline Leroy, *Predation success by a plant-ant indirectly favours the growth and fitness of its host myrmecophyte*, en révision, pour PLoS ONE
6. O. Roux, V. Rossi, R. Céréghino, A. Compin, J.M. Martin, A. Dejean, *How to coexist with fire ants : the roles of behaviour and cuticular compounds*, en révision pour Animal Behaviour
7. E. Marcon, I. Scotti, B. Hérault, V. Rossi, G. Lang, *Generalization of the partitioning of Shannon diversity*, soumis à PLoS ONE

### 2.2 Articles publiés dans des revues avec facteur d'impact

1. F. Mortier, V. Rossi, G. Guillot, S. Gourlet-Fleury, N. Picard, *Population dynamics of species-rich ecosystems : the mixture of matrix population models approach*, Methods in Ecology and Evolution, sous presse
2. Q. Molto, V. Rossi, L. Blanc, *Error propagation in biomass estimation in tropical forests*, Methods in Ecology and Evolution, sous presse
3. J.M. Thevenin, V. Rossi, M. Ducamp, P. Lachenaud, *Numerous clones resistant to Phytophthora palmivora in the "Guiana" genetic group of Theobroma cacao L*, PLoS ONE 7(7) : e40915 (2012)
4. F. Wagner, V. Rossi, C. Stahl, D. Bonal, B. Hérault, *Water availability is the main climate driver of neotropical tree growth*, PLoS ONE 7(4) : e34074 (2012)
5. A. Déjean, R. Céréghino, J.M. Carpenter, B. Corbara, B. Hérault, V. Rossi, M. Leponce, J. Orivel, D. Bonal, *Climate change impact on Neotropical social wasps*, PLoS ONE 6(6) : e27004 (2011)
6. B. Hérault, B. Bachelot, L. Poorter, V. Rossi, F. Bongers, J. Chave, C. E.T. Paine, F. Wagner and C. Baraloto, *Functional traits shape ontogenetic growth trajectories of rain forest tree species*, Journal of Ecology, 99(6) : 1431-1440 (2011)
7. F. Wagner, B. Hérault, C. Stahl, D. Bonal, V. Rossi, *Modeling water availability for trees in tropical forests*, Agricultural and Forest Meteorology, 151(9) : 1202-1213 (2011)
8. S. Gourlet-Fleury, V. Rossi, M. Rejou-Mechain, V. Freycon, A. Fayolle, L. Saint-André, G. Cornu, J. Gérard, J.M. Sarrailh, O. Flores, F. Baya, A. Billand, N. Fauvet, M. Gally, M. Henry, D. Hubert, A. Pasquier, N. Picard, *Environmental filtering of dense-wooded species controls above-ground biomass stored in African moist forests*, Journal of Ecology, 99(4) : 981-990 (2011)
9. A. Dejean, B. Corbara ; C. Leroy, J.H.C. Delabie, V. Rossi ; R. Céréghino, *Inherited biotic protection in a Neotropical pioneer plant*, PLoS ONE 6(3) : e18071 (2011)
10. H. Wernsdörfer, H. Caron, S. Gerber, G. Cornu, V. Rossi, F. Mortier, S. Gourlet-Fleury, *Relationships between demography and gene flow and their importance for the conservation of tree populations in tropical forests under selective felling regimes*, Conservation Genetics, vol. 12, n.1, 15-29 (2011)

11. A. Dejean, C. Leroy, B. Corbara, R. Céréghino, O. Roux, B. Hérault, V. Rossi, R. J. Guerrero, J. H. C. Delabie, J. Orivel, R. Boulay, *A temporary social parasite of tropical plant-ants improves the fitness of a myrmecophyte*, *Naturwissenschaften*, vol. 97, n.10, 925-934 (2010)
12. N. Picard, F. Mortier, V. Rossi, S. Gourlet-Fleury, *Clustering species using a model of population dynamics and aggregation theory*, *Ecological Modelling*, 221 : 152-160 (2010)
13. F. Campillo, V. Rossi, *Convolution Particle Filter for Parameter Estimation in General State-Space Models*, *IEEE Transactions on Aerospace and Electronic Systems*, vol.45, n.3, p1063-1072 (2009)
14. F. Campillo, R. Rakotozafy, V. Rossi, *Parallel and interacting Markov chain Monte Carlo algorithm*, *Mathematics and Computers in Simulation* 79 : 3424-3433 (2009)
15. O. Florès, V. Rossi, F. Mortier, *Bayesian Models of local density in tropical saplings : autocorrelation balances zero-inflation*, *Ecological Modelling*, 220 : 1797-1809 (2009)
16. H. Wernsdörfer, V. Rossi, G. Cornu, S. Oddou-Muratorio, S. Gourlet-Fleury, *Impact of uncertainty in tree mortality on the predictions of a tropical forest dynamics model*, *Ecological Modelling* 218 (3-4) : 290-306 (2008)
17. V. Rossi, J-P Vila, *Bayesian Selection of Multiresponse Nonlinear Regression Model*. *Statistics*, 42 :4,291-311 (2008)
18. O. Gimenez, V. Rossi, R. Choquet, C. Dehais, B. Doris, H. Varella, JP,R. Pradel, *State-space modelling of data on marked individuals*. *Ecological Modelling*. 206 : 431-438. (2007)
19. N. Hilgert, V. Rossi, J-P. Vila, V. Wagner, *Identification, estimation and control of uncertain dynamic systems : a nonparametric approach*. *Communications in Statistics : Theory and Methods*, vol. 36, n.14, p2509-2525 (2007).
20. V. Rossi, J-P. Vila, *Bayesian Multioutput Feedforward Neural Network Selection : A Conjugate Prior Approach*. *IEEE Transaction on Neural Networks*, vol.17, n.1, p35-47 (2006).
21. V. Rossi, J-P. Vila, *Approche non paramétrique du filtrage de système non linéaire à temps discret et à paramètres inconnus*. *C. R. Acad. Sci. Paris, Ser. I* 340, p759-764 (2005).

## 2.3 Articles publiés dans des revues à comité de lecture, sans facteur d'impact

1. F. Campillo, P. Cantet, R. Rakotozafy, V. Rossi, *Méthodes MCMC en interaction pour l'évaluation de ressources naturelles*. Revue ARIMA, 8 (spécial CARI'06) : 64-80.
2. F. Campillo, R. Rakotozafy, V. Rossi, *Computational probability modeling and Bayesian inference*. Revue ARIMA, 9 : 123-143.
3. V. Rossi, J-P. Vila, *Nonlinear filtering in discrete time : A particle convolution approach*. Ann. I.S.U.P., vol.50, n.3, p71-102
4. V. Rossi, J-P. Vila, *Filtrage de bioprocédé de dépollution. Approche par convolution particulière*. e-STA. vol. 2, n.1, (2005)

## 2.4 Chapitre de livre

1. M. Joannides, I. Larramendy-Valverde, V. Rossi, *Monte Carlo observer for a stochastic model of bioreactors*. In : Skiadas Christos H. (ed.). *Recent advances in stochastic modeling and data analysis*. Singapour : World Scientific Publishing [SG], p. 605-612, 2007

## 2.5 Actes de conférences

1. N. Hilgert, V. Rossi, J-P. Vila, *Identification, estimation and control of uncertain dynamic systems : a nonparametric approach*. International Symposium on Applied Stochastic Models and Data Analysis, Part VIII, p775-783, Brest (France), 2005
2. V. Rossi, J-P. Vila, *Radial Basis Network approach for non linear filtering in discrete time*. Proceedings of Int. Joint Conf. on Neural Networks, vol4, p2433- 2437, Portland (Oregon), 2003
3. V. Rossi, J-P. Vila, *Multioutput feedforward neural network selection : a bayesian approach*. Proceedings of Int. Joint Conf. on Neural Networks, vol1, p495- 500 Portland (Oregon), 2003

4. V. Rossi, J-P. Vila, *Filtrage non linéaire en temps discret par convolution de particules*. Actes des XXXVème Journées de Statistique, p823-826, Lyon, 2003
5. V. Rossi, J-P. Vila, *Critère de sélection de modèles multi-réponses par approche bayésienne*. Application aux réseaux de neurones. Actes des XXXI-Vème Journées de Statistique, p126-127, Bruxelles et Louvain-la-neuve, 2002.

## 2.6 Rapports de recherche

1. F. Campillo, V. Rossi, *Parallel and interacting Markov chains Monte Carlo method*. Research report n.6008, INRIA, 30p, 2006
2. F. Campillo, V. Rossi, *Convolution particle filters for parameter estimation in general state-space models*. Research report n.5939, INRIA, 28p, 2006
3. V. Rossi, *Filtrage non linéaire par noyaux de convolution. Application à un procédé de dépollution biologique*, Rapport de Thèse, Ecole Nationale Supérieure Agronomique de Montpellier, 279p, 2004
4. V. Rossi, J-P. Vila, *Bayesian selection of Multiresponse Nonlinear Regression Model*. Technical report 04-01, groupe biostatistique ENSAM/INRA/UM2 Montpellier, 2004
5. V. Rossi, J-P. Vila, *Bayesian multireponse nonlinear regression model selection : A conjugate prior approach*. Technical report 02-10, groupe biostatistique ENSAM/INRA/UM2 Montpellier, 2002
6. V. Rossi, *Critère de sélection de modèles multi-réponses par approche bayésienne. Application aux réseaux de neurones à couches*. Master 2 report, University Montpellier II, 2001

## 2.7 Communications dans des congrès internationaux

1. F. Wagner, M. Aubry-Kientz, A. L. Moore, Q. Molto, V. Rossi, B.Hérault, *Confronting satellite and field measurement data to improve the understanding of carbon uptake by tree growth in French Guiana*, Cayenne, 15ème Symposium SELPER L'Observation de la Terre pour un monde plus vert et plus solidaire, Cayenne, Guyane française, novembre 2012

2. P. Lachenaud, V. Rossi, JM Thevenin, M. Ducamp, F. Doaré, V. Condina, C. Hérail, F. Dedieu, *Evaluation de la résistance à Phytophthora palmivora et P. capsici des 186 clones de la « core collection » du groupe génétique « Guiana »*, 17ème conférence de l'Alliance des pays producteurs de cacao, Yaounde, Cameroun, octobre 2012
3. M. Aubry-Kientz, Ayotte-Trépanier C., Hérault B., Rossi V., *Towards trait-based mortality models for tropical forests ?* In Annual Meeting of the Association for Tropical Biology and Conservation, Bonito, Brazil, June 2012
4. F. Wagner, A. Moore, V. Rossi, B. Hérault, *Which climate indices are relevant for predicting the response of tropical forests to future climate scenarios ?* In Annual Meeting of the Association for Tropical Biology and Conservation, Bonito, Brazil, June 2012
5. P. Sist, Mazzei L., Blanc L., Ruschel A., Rossi V., Kanshiro M., *Long term Impact of logging on carbon storage and tree diversity in the Amazon*, In Annual Meeting of the Association for Tropical Biology and Conservation, Bonito, Brazil, June 2012
6. F. Wagner, B. Hérault, C. Stahl, D. Bonal, V. Rossi, *Assessing the impact of water availability on growth of neotropical trees*. In International Joint Meeting of the ATBC and SCB Africa, Arusha, Tanzania, 12-16 June, 2011
7. Q. Molto, V. Rossi, L. Blanc, *Error propagation through a bayesian network for biomass estimation in neotropical forests*. International Biometric Conference, Floripa, Brazil, 5-10 December 2010.
8. F. Wagner, B. Hérault, D. Bonal, V. Rossi, *Modeling soil water availability for neotropical forest trees*. International Biometric Conference, Floripa, Brazil, 5-10 December 2010.
9. F. Mortier, V. Rossi, N. Picard, S. Gourlet-Fleury, *Unsupervised classification of species groups based on mixture matrix population models*. International Biometric Conference, 13-18 July, 2008, Dublin, Irlande.
10. V. Rossi, N. Desassis, F. Campillo, *Individual-based modelling of spatio-temporal of forest dynamics*. Meeting of the Statistical Society of Canada and the Société Française de Statistique, May 25 to 29, 2008, Ottawa, Canada.
11. M.I. Peyre, S. Desvaux, T. Phan Dang, V. Rossi, J.F. Renard, Vu Dinh Tuan, F. Roger, *Financial evaluation of vaccination strategies against highly pathogenic avian influenza*. International Conference on Avian Influenza, Bangkok, Thailand, January 23-25, 2008.

12. F. Campillo, R. Rakotozafy, V. Rossi, *Bayesian numerical inference for hidden Markov models*. Colloque international de Statistique Appliquée pour le Développement en Afrique - SADA'07, 26 février - 02 mars 2007, Cotonou, Bénin
13. F. Campillo, R. Rakotozafy, V. Rossi, *Bayesian numerical inference for markovian models. Application to tropical forest dynamics*. International Conference on Approximation Methods and Numerical Modelling in Environment and Natural Resources, July 11-13, 2007, Granada (Spain).
14. Mortier F., Rossi V., Picard N., Gourlet-Fleury S., *Unsupervised classification of species groups based on mixture matrix population models*. In Conférence META 2007, Kourou, Guyane.
15. M. Joannides, I. Larramendy-Valverde, V. Rossi, *Monte Carlo observer for a stochastic model of bioreactors*. In International conference on Applied stochastic models and data analysis. 12, 2007-05-29/2007-06-01, Chania, Grèce.
16. F. Campillo, V. Rossi, *Convolution particle filtering for parameter estimation in general state-space models*. The 45th IEEE Conference on Decision and Control, San Diego, december 2006.
17. F. Campillo, R. Rakotozafy, V. Rossi, *Analyse bayésienne de modèles markoviens d'évolution de ressources naturelle*. Colloque Africain sur la Recherche en Informatique, Cotonou Bénin, november 2006
18. N. Hilgert, V. Rossi, JP Vila, *Identification, estimation and control of uncertain dynamic systems : a nonparametric approach*. International Symposium on Applied Stochastic Models and Data Analysis, Brest (France)
19. V. Rossi, J-P. Vila, *Filtrage de bioprocédé de dépollution. Approche par convolution particulière*. Conférence Francophone Internationale d'Automatique, n.125, Douz (Tunisia), 2004
20. V. Rossi, J-P. Vila, *Radial Basis Network approach for non linear filtering in discrete time*. Int. Joint Conf. on Neural Networks, Portland (Oregon), 2003
21. V. Rossi, J-P. Vila, *Multioutput feedforward neural network selection : a bayesian approach*. Int. Joint Conf. on Neural Networks, Portland (Oregon), 2003

## 2.8 Communications dans des congrès nationaux

1. M. Aubry-Kientz, B. Hérault, C. Ayotte-Trepanier, V. Rossi, *Modélisation de la mortalité des arbres en forêt tropicale avec covariables incertaines*, 7ème Rencontres Statistiques de Rochebrune, 1-6 avril 2012
2. F. Mortier, P. Chagneau, MP Etienne, N. Picard, C. Piou, V. Rossi, *Modélisation bayésienne hiérarchique pour l'écologie et la recherche environnementale*. In : SFDS, Société française de statistique. 43èmes Journées de statistique, Gammarth, Tunisie, 23-27 mai 2011
3. Q. Molto, V. Rossi, L. Blanc, *Error propagation for biomass estimation in neotropical forest*. In Premier Colloque national d'écologie scientifique, Montpellier, 2-4 septembre 2010.
4. H. Wernsdörfer, V. Rossi, G. Cornu, S. Oddou-Muratorio, H. Caron, S. Gerber, S. Gourlet-Fleury, *Modelling the impact of selective felling on the demography and genetic diversity of *Dicorynia guianensis*, the major timber species in the tropical rainforest of French Guiana*. 6th IUFRO Workshop of "uneven-aged silviculture" in Shizuoka, october 2008.
5. F. Campillo, N. Desassis, V. Rossi, *Modélisation spatialement explicite d'une dynamique forestière et inférence*, Journées MAS de la SMAI, Rennes, aout 2008
6. N. Picard, S. Gourlet-Fleury, F. Mortier, V. Rossi, G. Cornu, P. Chagneau, *Modèles matriciels pour la gestion de population naturelle structurée en taille, Application à la gestion forestière en Afrique centrale*, Colloque INRA Modélisation pour les ressources naturelles Montpellier, juin 2008
7. H. Wernsdörfer, G. Cornu, V. Rossi, S. Gourlet-Fleury, S. Gerber, *Evaluation et amélioration d'un modèle arbre dépendant des distances utilisé pour simuler le flux de gènes. Cas de l'Angélique ( *Dicorynia guinensis*) en Guyane Française*, 10ème Journées du réseau Croissance, Améliorations, Qualité des bois (CAQ10), Bordeaux, mars 2007
8. F. Campillo, V. Rossi, *Analyse bayésienne de modèles d'évolution de ressources naturelles*. Journées MAS de la SMAI, Lille, september 2006
9. V. Rossi, *Introduction d'une sélection pour éviter la divergence des filtres particuliers à convolution*. Colloque Statistique des processus - Applications au traitement du signal et de l'image, Institut de Mathématiques Appliquées à Angers, 2005



10. V. Rossi, *Utilisation du filtrage particulaire en dynamique des populations et sur bio-procédés*. Participation en qualité d'invité aux Journées de Statistique et applications en biologie, Université de Rennes 2, septembre 2005
11. V. Rossi, *Estimation particulaire de paramètres dans les systèmes dynamiques cachés*. Premières Rencontres des Jeunes Statisticiens, Aussois, 2005
12. V. Rossi, J-P. Vila, *Filtrage non linéaire par convolution pour modèles incertains*. XXXVIème Journées de Statistique, Montpellier, 2004
13. V. Rossi, J-P. Vila, *Filtrage non linéaire en temps discret par convolution de particules*. XXXVème Journées de Statistique, Lyon, 2003
14. V. Rossi, *Utilisation des réseaux de neurones pour la régression multivariée. Sélection de la structure du réseau par un critère bayésien*. Participation en qualité d'invité à la Journée Réseaux de neurones et applications du ClubSAS Stat, ENSAE Paris, 2002
15. V. Rossi, J-P. Vila, *Critère de sélection de modèles multi-réponses par approche bayésienne. Application aux réseaux de neurones*. XXXIVème Journées de Statistique, Bruxelles et Louvain-la-neuve, 2002
16. V. Rossi, J-P. Vila, *Utilisation des réseaux de neurones à bases radiales pour le filtrage non linéaire en temps discret*. XIème Journées de Neurosciences et Science pour l'Ingénieur, La londe les Maures, 2002

## 3 Encadrement d'étudiants

### 3.1 Thèses

**Mélaine Aubry-Keintz** (depuis octobre 2011). Sujet : Modéliser la dynamique des peuplements forestiers guyanais pour comprendre les conséquences des changements climatiques sur le futur de la forêt tropicale. Thèse de doctorat de l'Université de Antilles et de la Guyane. Co-encadrants Bruno Hérault (directeur) et Vivien Rossi.

#### *Publication*

- M. Aubry-Kientz, B. Hérault, C. Ayotte-Trepanier, C. Baraloto, V. Rossi, *Towards trait-based mortality models for tropical forests ?*, en révision pour PLoS ONE

#### *Communications*

- Aubry-Kientz M., C. Ayotte-Trépanier, B. Hérault, V. Rossi, *Towards trait-based mortality models for tropical forests ?* In Annual Meeting of the Association for Tropical Biology and Conservation, Bonito, Brazil, June 2012
- Aubry-Kientz M., Hérault B., Ayotte-Trepanier C., Rossi V., *Modélisation de la mortalité des arbres en forêt tropicale avec covariables incertaines*, 7ème Rencontres Statistiques de Rochebrune, 1-6 avril 2012

**Quentin Molto** (octobre 2011 - décembre 2013). Sujet : Estimation de biomasse en forêt tropicale humide, propagation des incertitudes dans la modélisation de la distribution spatiale de la biomasse en Guyane française. Thèse de doctorat de l'Université de Antilles et de la Guyane. Co-encadrants Bruno Hérault (directeur), Lilian Blanc et Vivien Rossi.

#### *Publications*

- Q. Molto, V. Rossi, L. Blanc, *Error propagation in biomass estimation in tropical forests*, *Methods in Ecology and Evolution*, sous presse

- Q. Molto, B. Hérault, O. Brunaux, G. Coste, V. Freycon, S. Guitet, L. Blanc, V. Rossi, *Mapping Biomass of tropical forests : uncertainty propagation with a model-base approach*, en préparation
- Q. Molto, B. Hérault, JJ Boreux, M. Daullet, A. Rousteau, V. Rossi, *Predicting tree height for biomass estimates in tropical forest*, en préparation

#### *Communications*

- Molto Q., Rossi V., Blanc L., *Error propagation for biomass estimation in neotropical forest*. In Premier Colloque national d'écologie scientifique, Montpellier, 2-4 septembre 2010.
- Molto Q., Rossi V., Blanc L., *Error propagation through a bayesian network for biomass estimation in neotropical forests*. International Biometric Conference, Floripa, Brazil, 5-10 December 2010.

**Fabien Wagner** (octobre 2008 - décembre 2011). Sujet : La réponse des forêts tropicales humides aux variations climatiques. Evolution de la structure et de la dynamique des peuplements forestiers guyanais. Thèse de doctorat de l'Université de Antilles et de la Guyane. Co-encadrants Bruno Hérault (directeur) et Vivien Rossi.

#### *Publications*

- F. Wagner, V. Rossi, C. Baraloto, D. Bonal, C. Stahl, B. Hérault, *Functional traits explain tropical tree growth response to climate*, soumis à Functional Ecology
- F. Wagner, V. Rossi, C. Stahl, D. Bonal, B. Hérault, *Water availability is the main climate driver of neotropical tree growth*, PLoS ONE 7(4) :e34074 (2012)
- F. Wagner, B. Hérault, C. Stahl, D. Bonal, V. Rossi, *Modeling water availability for trees in tropical forests*, Agricultural and Forest Meteorology, 151(9) : 1202-1213 (2011)

#### *Communications*

- F. Wagner, B. Hérault, C. Stahl, D. Bonal, V. Rossi, *Assessing the impact of water availability on growth of neotropical trees*. In International Joint Meeting of the ATBC and SCB Africa, Arusha, Tanzania, 12-16 June, 2011
- F. Wagner, B. Hérault, D. Bonal, V. Rossi, *Modeling soil water availability for neotropical forest trees*. International Biometric Conference, Floripa, Brazil, 5-10 December 2010.

Fabien Wagner a continué en post-doctorat dans l'UMR Ecofog sur le projet Clim-For.

## 3.2 Post-Doctorats

**Fabien Wagner** (mai 2012 - juin 2013). Sujets : modélisation des impacts du changement climatique sur la biodiversité de la forêt tropicale de la Guyane française. Post-doctorat dans l'UMR Ecofog dans le cadre du projet ClimFor. Co-encadrants : Bruno Hérault et Vivien Rossi.

### *Publications*

- F. Wagner, A. Moore, V. Rossi, B. Hérault, *The intra-annual effect of climate on tree growth in tropical forest, a Meta analysis*, en préparation
- F. Wagner, M. Aubry-Kientz, C. Stahl, D. Bonal, Q. Molto, A. Moore, L. Linguet, Y. Goulamoussene, V. Rossi, B. Hérault, *Confronting satellite and field measurement data to improve the understanding of carbon uptake by tree growth in French Guiana*, en préparation

### *Communications*

- F. Wagner, M. Aubry-Kientz, A. L. Moore, Q. Molto, V. Rossi, B. Hérault, *Confronting satellite and field measurement data to improve the understanding of carbon uptake by tree growth in French Guiana*, Cayenne, 15ème Symposium SELPER L'Observation de la Terre pour un monde plus vert et plus solidaire, Cayenne, Guyane française, novembre 2012
- F. Wagner, A. Moore, V. Rossi, B. Hérault, *Which climate indices are relevant for predicting the response of tropical forests to future climate scenarios ?* In Annual Meeting of the Association for Tropical Biology and Conservation, Bonito, Brazil, June 2012

**Alana Moore** (novembre 2011 - novembre 2012). Sujets : Evaluating risks of climate change on tropical forest ecosystem services in French Guiana regarding the protected area network. Post-doctorat dans l'UMR Ecofog dans le cadre du cofinancement Université de Melbourne et projet GuyaSim. Co-encadrants : Bruno Hérault et Vivien Rossi.

*Publications et Communications : cf. ci-dessus*

**Holger Wernsdörfer** (janvier 2007 - janvier 2009). Sujets : Etude de l'impact de la mortalité des juvéniles sur la dynamique forestière par l'intermédiaire d'une analyse de sensibilité du modèle SELVA et étude de l'impact de l'exploitation sur le flux de gènes dans les forêts tropicales. Post-doctorat dans l'UR Dynamiques des forêts naturelles du Cirad. Co-encadrants : Sylvie Gourlet-Fleury et Vivien Rossi.

### *Publications*

- H. Wernsdörfer, H. Caron, S. Gerber, G. Cornu, V. Rossi, F. Mortier, S. Gourlet-Fleury, *Relationships between demography and gene flow and their importance for the conservation of tree populations in tropical forests under selective felling regimes*, Conservation Genetics, vol. 12, n.1, 15-29 (2011)
- H. Wernsdörfer, V. Rossi, G. Cornu, S. Oddou-Muratorio, S. Gourlet-Fleury, *Impact of uncertainty in tree mortality on the predictions of a tropical forest dynamics model*, Ecological Modelling 218 (3-4) : 290-306 (2008)

### *Communications*

- H. Wernsdörfer, V. Rossi, G. Cornu, S. Oddou-Muratorio, H. Caron, S. Gerber, S. Gourlet-Fleury, *Modelling the impact of selective felling on the demography and genetic diversity of Dicorynia guianensis, the major timber species in the tropical rainforest of French Guiana*. 6th IUFRO Workshop of "uneven-aged silviculture" in Shizuoka, october 2008.
- H. Wernsdörfer, G. Cornu, V. Rossi, S. Gourlet-Fleury, S. Gerber, *Evaluation et amélioration d'un modèle arbre dépendant des distances utilisé pour simuler le flux de gènes. Cas de l'Angélique ( Dicorynia guianensis) en Guyane Française*, 10ème Journées du réseau Croissance, Améliorations, Qualité des bois (CAQ10), Bordeaux, mars 2007

Holger Wernsdörfer a ensuite été recruté comme Maître de conférence par Agro-ParisTech dans l'UMR Lerfob à Nancy.

**Nicolas Desassis** (avril 2007 - mars 2008). Sujet : Développement d'un modèle de dynamique forestière arbre centré, opérationnel en temps continu, intégrant une compétition de type densité dépendance pour un peuplement mono-spécifique. Post-doctorat dans l'équipe MERE de l'Inria, dans le cadre de l'ARC MIRC. Co-encadrants Fabien Campillo et Vivien Rossi.

### *Communications*

- V. Rossi , N. Desassis, F. Campillo, *Individual-based modelling of spatiotemporal of forest dynamics*. Meeting of the Statistical Society of Canada and the Société Française de Statistique, May 25 to 29, 2008, Ottawa, Canada.
- F. Campillo, N. Desassis, V. Rossi, *Modélisation spatialement explicite d'une dynamique forestière et inférence*, Journées MAS de la SMAI, Rennes, aout 2008

Nicolas Desassis a été recruté, à l'issu de ce post-doc, comme enseignant-chercheur par MINES ParisTech à Fontainebleau.

### 3.3 Stages de DEA et Master 2

**Charles Ayotte-Trepanier** (5 mois en 2011). Sujet : Modélisation de la mortalité des arbres à l'échelle de la communauté. Master 2 de Biostatistique, Université Montpellier 2. Co-encadrants : Bruno Hérault et Vivien Rossi.

**Christelle Apha Niango** (6 mois en 2010). Sujet : Segmentation du type forestier sur la base de levés LIDAR. M2 Mathématiques appliquées : Ingénieur Statistique, Université Grenoble 1. Co-encadrants : Vivien Rossi et Grégoire Vincent

**Clara Landry** (5 mois en 2010). Sujet : Modélisation bayésienne de la dynamique d'un peuplement forestier agrégé avec prise en compte de la densité dépendance. Application à la prédiction de reconstitution de massifs forestiers guyanais exploités. Master 2 de Biostatistique, Université Montpellier 2. Co-encadrants : Lilian Blanc et Vivien Rossi.

Clara Landry a ensuite fait un volontariat de service civique d'un an au Cirad en Guadeloupe. Puis, en octobre 2011, elle y a démarré une thèse sur la modélisation d'épidémies foliaires de culture pérenne tropicale à différentes échelles spatiales appliquée au pilotage des variétés.

**Fabien Rollot** (5 mois en 2009). Sujet : Modélisation de la croissance d'essences forestières tropicales prenant en compte la variabilité interspécifique. Master 2 de Biostatistique, Université Montpellier 2. Co-encadrants : Frédéric Mortier et Vivien Rossi

**Nabil Zougab** (5 mois en 2008). Sujet : Développement d'un modèle de génétique quantitative dynamique. Prise en compte des compétitions inter-individuelles. Master 2 de Biostatistique, Université Montpellier 2. Co-encadrants : Frédéric Mortier et Vivien Rossi.

Nabil Zougab est ensuite rentré en Algérie pour occuper un poste d'assistant à l'Université de Bejaia dans le laboratoire de Modélisation et Optimisation des Systèmes (LAMOS). Il s'est également lancé dans une thèse.

**Marie Denis** (5 mois en 2007). Sujet : Analyse bayésienne de modèles d'évolution de ressources naturelles. Master 2 de Biostatistique, Université Montpellier 2. Co-encadrants : Fabien Campillo et Vivien Rossi

Marie Denis a poursuivi en thèse à l'Université de Montpellier 1 sous la direction de Nicolas Molinari. Elle a soutenu son mémoire intitulé *Méthodes de modélisation bayésienne et applications en recherche clinique* en juillet 2007. Elle a ensuite été recrutée comme chercheuse au Cirad dans l'UMR AGAP.

**Marie Cheminat** (5 mois en 2007). Sujet : Filtrage particulaire et calcul du facteur de Bayes : application à la modélisation intégrée et à inférence en dynamique des populations. Master 2 de Biostatistique, Université Montpellier 2. Co-encadrants : Vivien Rossi et Jean-Pierre Vila

**Hubert Varella** (5 mois en 2006). Sujet : Modélisation intégrée en dynamique des populations et application du filtre particulaire avec noyaux de convolution. Master 2 de Biostatistique, Université Montpellier 2. Co-encadrants : Olivier Gimenez et Vivien Rossi

Hubert Varella a poursuivi en thèse l'Inra d'Avignon sous la direction de Martine Guérif. Il est à présent chercheur à Météo-France à Toulouse.

**Blaise Doris** (5 mois en 2005). Sujet : Modélisation intégrée en dynamique des populations : application du filtrage particulaire. Master 2 de Biostatistique, Université Montpellier 2. Co-encadrants : Vivien Rossi et Jean-Pierre Vila

### 3.4 Stages d'élèves ingénieurs

**Gustave Coste** (6 mois en 2011). Sujet : Extrapolation spatiale à l'échelle du massif forestier guyanais des estimations de biomasse aérienne. Césure élève ingénieur 3ème année Montpellier SupAgro. Co-encadrants : Quentin Molto et Vivien Rossi.

**Philippe Cantet** (5 mois en 2006). Sujet : Inférence Bayésienne par Méthodes de Monte Carlo pour des Modèles d'Évolution de ressources Renouvelables. 2ème année ingénieur ISTIL Lyon filière MCS. Co-encadrants : Fabien Campillo et Vivien Rossi.

Philippe Cantet a ensuite poursuivi en thèse à l'Université de Montpellier 2 sous la direction de Jean-Noël Bacro. Il a soutenu son mémoire intitulé *Impacts du changement climatique sur les pluies extrêmes par l'utilisation d'un générateur stochastique de pluies*, en décembre 2009.

### 3.5 Stages de Master 1

**Kejun Zou** (4 mois en 2012). Sujet : Effets des variations spatiales et temporelles du climat sur la croissance de *Dicorynia guianensis* en Guyane française. Master 1 Science et Technologies du Vivant et de l'Environnement, AgroParisTech. Co-encadrants : Vivien Rossi et Fabien Wagner.

**Rukiye Mutlu** (100h en 2009). Sujet : L'étude de la croissance et de la mortalité des arbres en République Centrafricaine. Master 1 Mathématique, Université Montpellier 2. Co-encadrants : Dakis Ouedraogo et Vivien Rossi

**Rémi Tylski** (4 mois en 2009). Sujet : Développement Web et Client lourd pour un calculateur d'évaluation des stratégies de vaccination en épidémiologie animale. Master 1 Informatique, Université du Littoral Côte d'Opale. Co-encadrants : Marisa Peyre et Vivien Rossi

### 3.6 Comités de thèse

**Dakis Ouedraogo** (soutenue en décembre 2011). Sujet : Prédiction de la dynamique forestière à l'aide d'un modèle matriciel qui incorpore la variabilité de la réponse des espèces à l'environnement : application dans une forêt tropicale humide semi-décidue d'Afrique centrale. Doctorat de l'Université de Montpellier 2, Ecole Doctorale SIBAGHE. Directeur de thèse : Jean-Dominique Lebreton

**Marie Denis** (soutenue en juillet 2010). Sujet : Méthodes de modélisation bayésienne et applications en recherche clinique. Doctorat de l'Université de Montpellier 1, Ecole Doctorale I2S. Directeur de thèse : Nicolas Molinari.

**Pierre Chagneau** (soutenue en décembre 2009) Sujet : Modélisation hiérarchique bayésienne pour la prédiction multivariée de processus spatiaux non-gaussiens et processus ponctuels hétérogènes d'intensité liée à une variable prédite. Application à la prédiction de la régénération en forêt tropicale humide. Doctorat de l'Université de Montpellier 2, Ecole Doctorale I2S. Directeur de thèse : Jean-Noël Bacro.

**Ghislain Vieilledent** (soutenue en février 2009). Sujet : Structurer l'incertitude et la variabilité dans les modèles de dynamique forestière. Application à la coexistence du Sapin et de l'Epicéa en forêt de montagne. Doctorat de l'AgroParisTech-Engref, Ecole Doctorale RP2E. Directeur de thèse : Jean-François Dhôte.



## **Deuxième partie**

### **Travaux de recherche**



# 1 Introduction

Etant particulièrement attiré par les mathématiques appliquées aux sciences du vivant, je me suis orienté, après un cursus universitaire classique en mathématiques, vers un DEA de Biostatistique. J'ai ensuite choisi de faire une thèse à l'Inra dans le département Mathématique et Informatique Appliqués car au-delà de sa dimension méthodologique le sujet permettait d'avancer concrètement sur des problématiques de dépollution d'eaux usées. En plus de mon activité de recherche durant ma thèse et l'année suivante en tant qu'attaché temporaire d'enseignement et de recherche, j'ai enseigné la statistique dans des universités et dans écoles d'ingénieurs.

Puis, j'ai poursuivi ma formation à la recherche en effectuant un séjour post-doctoral dans l'équipe Aspi de l'Inria Rennes reconnue pour ses compétences méthodologiques dans le domaine de recherche de ma thèse. Bien que ses champs d'applications soient essentiellement les technologies de pointe, j'ai pu continuer à travailler sur des questions biologiques qui me tenaient à cœur comme la gestion de ressources naturelles. Mon recrutement au Cirad l'année suivante a constitué l'aboutissement de mon projet professionnel commencé cinq ans plus tôt.

J'ai intégré le Cirad en tant que modélisateur biostatisticien en 2007 dans l'UR Dynamique des Forêts Naturelles. Ma principale mission était de modéliser les différents processus de la dynamique forestière pour être en mesure de tester statistiquement des hypothèses écologiques. Depuis 2009, je suis affecté dans l'UMR Ecofog en Guyane Française. Cette mobilité m'a permis de me rapprocher de mon objet d'étude et de développer mon réseau professionnel. En tant que référent en modélisation et en statistique dans mes unités, j'ai travaillé avec des collègues forestiers, écologues ou épidémiologistes. Mon apport en modélisation a permis de faire des avancées sur des problématiques de diverses disciplines, concrétisées par des publications.

Mes activités scientifiques ont ainsi évolué naturellement durant mon parcours. Mes travaux de recherche initiaux concernaient surtout des questions de méthodo-

logie en mathématiques appliquées et statistique. A présent, mes questions de recherche portent essentiellement sur le développement de modèles pertinents pour des problématiques concernant les forêts tropicales. Ainsi les connaissances acquises grâce à mes travaux initiaux me servent de base pour mes activités actuelles.

La présentation de mes travaux de recherche reprend ces deux volets de mes activités. La première partie est une synthèse de mes activités de recherche méthodologique en mathématiques appliquées qui sont généralement les plus anciennes. La deuxième partie est une synthèse de mes activités de recherche en modélisation autour de la forêt tropicale. Ces travaux étant le fruit de collaborations pluridisciplinaires, je décris essentiellement ma contribution en modélisation, sans rentrer dans les détails sur la portée écologique de nos résultats. La dernière partie présente les thèmes et questions de recherche auxquels j'envisage de me consacrer les prochaines années.

## **2 Méthodologie de modélisation statistique**

Mon activité de modélisation consiste à formaliser des concepts et savoirs écologiques dans des modèles cohérents avec les données à disposition et les hypothèses écologiques à tester. Pour répondre à ces besoins, le recours à la statistique bayésienne est efficace. Le formalisme bayésien hiérarchique est à présent largement utilisé en écologie (Clark & Gelfand 2006). Il offre en effet une grande souplesse dans la conception des modèles et permet d'intégrer différents types de données et d'informations pour leur inférence. Cependant, la modélisation bayésienne peut être difficile à mettre en œuvre concrètement car elle pose de nombreuses questions ouvertes de recherche pluridisciplinaire. Selon la nature du modèle construit il peut être nécessaire de développer et d'implémenter des algorithmes originaux pour inférer le modèle, notamment en présence d'informations multiples et de non-linéarités. Aussi, bien que la spécificité de chaque système naturel étudié interdise une adaptation immédiate, les développements réalisés comporte une partie transposable à d'autres systèmes.

Ainsi les travaux méthodologiques présentés dans ce chapitre ne sont pas spécifiques à une domaine d'application particulier. Ils sont applicables à des modèles de systèmes dynamiques et des modèles statistiques. Mais comme expliquer ci-dessus, pour les mettre en œuvre sur un objet d'étude donné, par exemple les forêts tropicales, un travail d'adaptation est nécessaire.

### **2.1 Inférence de modèles de systèmes dynamiques**

L'inférence d'un modèle désigne l'ensemble des actions visant à confronter le modèle aux données. En particulier, cela désigne l'estimation des paramètres et variables inconnus ou les tests que l'on peut faire sur les paramètres.

La prise en compte de données dans les modèles est aussi désignée par l'assimilation de données. Les méthodes de Monte-Carlo séquentielles (SMC) (cf. [Doucet et al. \(2001\)](#)) sur lesquelles j'ai travaillé dans ma thèse demeurent la seule alternative pour l'assimilation des données dans les modèles généraux. En effet, les techniques d'assimilation de données sont divisées en deux classes : les méthodes variationnelles et les méthodes séquentielles. Les méthodes variationnelles d'assimilation de données reposent sur la théorie du contrôle optimal : les paramètres du modèle sont obtenus par optimisation d'une fonction de coût en supposant que le modèle est exact. Cette approche est d'ailleurs très utilisée en météorologie ([Daley, 1991](#)). Les méthodes séquentielles se placent dans un contexte probabiliste. Les estimations sont réalisées séquentiellement par propagation de l'information au fil du temps. Ces techniques s'apparentent ainsi au filtrage. Mais les hypothèses requises par le filtre de Kalman (modèles linéaires, bruits gaussiens) ne sont généralement pas vérifiées dans la pratique. Il faut alors avoir recours à une méthode de Monte Carlo.

Initialement, les méthodes de Monte Carlo séquentielles classiques ne semblaient pas utilisables pour des grandes dimensions. Mais sur un problème réel en océanographie ([Bertino et al., 2003](#)) et en météorologie ([Van Leeuwen, 2003](#)) ont obtenus des résultats prometteurs en s'appuyant sur des méthodes de Monte-Carlo séquentielles avec ré-échantillonnage.

Or dans le contexte des systèmes naturels, il est rarement raisonnable de supposé le modèle exact. La nature multiple des informations, la présence d'erreur de mesure sur les données, ainsi que la prise en compte de stochasticité dans le modèle, requiert l'utilisation de modèles probabilistes, du type des modèles à espace d'états décrits dans la section suivante.

Ainsi, l'inférence bayésienne de modèles de systèmes naturels, en plus d'autoriser une formalisation plus réaliste, soulève donc des questions méthodologiques très pertinentes.

### 2.1.1 Modèles de Markov cachés ou modèles à espace d'état

Le cadre de travail des modèles à espace d'état ou modèles de Markov cachés est défini par un processus stochastique  $\{(X_t, Y_t); t \in \mathbb{N}\}$  où à l'instant  $t$ , les  $X_t \in \mathbb{R}^d$ , sont les variables d'état cachées et les  $Y_t \in \mathbb{R}^q$ , sont les variables observées. Le processus d'état est supposé markovien :

$$Loi(X_t | X_{t-1}, \dots, X_0) = Loi(X_t | X_{t-1}).$$

Les observations sont supposées indépendantes conditionnellement aux états :

$$Loi(Y_t, \dots, Y_0 | X_t, \dots, X_0) = \prod_{i=0}^t Loi(Y_i | X_i).$$

Un tel processus est schématisé par le graphe :

$$\begin{array}{ccccccc} X_1 & \longrightarrow & X_2 & \longrightarrow & \dots & \longrightarrow & X_t & \longrightarrow & \dots \\ \downarrow & & \downarrow & & & & \downarrow & & \\ Y_1 & & Y_2 & & & & Y_t & & \end{array}$$

Dans ce contexte, l'objectif est de fournir, dynamiquement dans le temps  $t$ , de l'information sur l'état du système,  $X_t$ , en utilisant toutes les données à disposition  $y_0, \dots, y_t$  à l'instant  $t$  et en s'appuyant sur la connaissance du système (i.e.  $Loi(X_t | X_{t-1})$  et  $Loi(Y_t | X_t)$  connues). Ce problème relève de la théorie mathématique du filtrage. La solution théorique du problème constitue le filtre optimal :  $Loi(X_t | Y_0 = y_0, \dots, Y_t = y_t)$ , la loi conditionnelle de l'état  $X_t$  sachant toutes les observations jusqu'à l'instant  $t$ ,  $y_0, \dots, y_t$ . Aussi dans la communauté du filtrage, tout algorithme récurrent visant à estimer le filtre optimal est traditionnellement désigné par filtre.

### Estimation de l'état dans les modèles de Markov cachés

Il existe de nombreux problèmes pratiques qui relèvent de la théorie du filtrage. Notamment, toutes les situations pour lesquelles il est difficile, voire impossible, de mesurer l'état  $X_t$  d'un système dynamique connu et où il faut alors se contenter d'observations bruitées  $Y_t$ . De manière plus pragmatique, tous les phénomènes physiques, chimiques ou biologiques modélisables par un système de la forme (2.1) relèvent de la théorie du filtrage en temps discret.

$$\begin{cases} x_t = f_t(x_{t-1}, \varepsilon_t) \\ y_t = h_t(x_t, \eta_t) \end{cases} \quad (2.1)$$

Où  $x_t \in \mathbb{R}^d$  sont les variables d'état cachées,  $y_t \in \mathbb{R}^q$  sont les variables observées,  $\varepsilon_t$  sont les bruits du modèle,  $\eta_t$  sont les bruits du modèle d'observation. Les fonctions  $f_t, h_t$  sont boréliennes et non nécessairement homogènes dans le temps. Les modèles de Markov cachés présentés sous la forme (2.1) sont aussi appelés modèles à espace d'états dans la littérature. Par ailleurs la discrétisation d'équations différentielles stochastiques conduit aussi à ce type de modèles.

Dans le cas où les fonctions  $f_t$  et  $h_t$  sont linéaires et les bruits blancs gaussiens additifs, le problème a été résolu par [Kalman \(1960\)](#). Le cas général des systèmes non linéaires constitue le cadre de recherche privilégié en filtrage, d'autant qu'il n'existe pas, le plus souvent, de solution de dimension finie ([Chaleyat-Maurel and Michel, 1984](#)).

Différentes approches ont été proposées pour estimer le filtre optimal, elles souffrent généralement de temps de calculs excessivement élevés lorsque la dimension des variables d'états dépasse quelques unités.

Dans ce contexte, l'intérêt croissant porté aux récentes méthodes de Monte Carlo, les filtres particulaires et leurs variantes, est naturel puisque ces techniques sont théoriquement insensibles à la dimension des variables. Elles visent à estimer la distribution du filtre optimal par des particules simulées à partir du modèle (2.1), dont l'évolution est liée aux observations. Ces approches sont intéressantes car elles s'appuient sur des résultats théoriques ([Del Moral, 1995](#)) et leurs performances surpassent généralement celles de leurs concurrents. Cependant, la mise en pratique des filtres particulaires requiert un bruit additif sur le modèle d'observation : ainsi peut-on évaluer la vraisemblance d'une particule et bâtir une mesure empirique. Ces méthodes particulaires sont alors défaillantes si les observations sont faiblement bruitées et deviennent impraticables si les observations ne sont pas bruitées.

Mes travaux initiaux en filtrage m'ont permis au final de proposer dans ma thèse quatre algorithmes pour construire des estimateurs de la densité du filtre optimal. J'ai établi les propriétés de convergence de ces filtres vers le filtre optimal, sous des conditions classiques sur le noyau et sur la largeur de la fenêtre. J'ai notamment développé un filtre par convolution avec ré-échantillonnage inspiré des filtres particulaires avec interactions introduits par ([Del Moral, 1995](#)). Pour ces derniers, une sélection, de type algorithme génétique, régule l'évolution des particules : les plus vraisemblables par rapport aux observations ont plus de chance de continuer à évoluer que les moins vraisemblables. Cette approche conserve donc naturellement les particules dans la zone de l'espace intéressante au vu des observations, améliorant ainsi la qualité du filtre. Ces résultats ont été présentés des conférences internationales et dans des articles ([Rossi and Vila, 2005b, 2006b](#)).



## Estimation de paramètres dans les modèles de Markov cachés

D'un point de vue pratique, il est très pertinent en filtrage de chercher à prendre en compte une méconnaissance sur le modèle utilisé. On considère le cas où l'incertitude est représentée par des paramètres inconnus  $\theta \in \mathbb{R}^p$  :  $Loi(X_t|X_t, \theta)$  et  $Loi(Y_t|X_t, \theta)$ . En terme de modèle à espace d'états, le système dynamique plus général, ci-dessous, nous place dans un tel contexte :

$$\begin{cases} x_t &= f_t(x_{t-1}, \theta, \varepsilon_t) \\ y_t &= h_t(x_t, \theta, \eta_t) \end{cases} \quad (2.2)$$

Les paramètres  $\theta$  sont donc à estimer parallèlement à l'état  $x_t$ . Durant ma thèse, j'ai exploré différentes approches pour traiter ce problème. Tout d'abord en adaptant des techniques statistiques classiques d'estimation de paramètres : les moindres carrés conditionnels et le maximum de vraisemblance. Ces estimateurs n'étant pas calculables en pratique, j'ai proposé une adaptation des algorithmes de mes filtres pour les approximer. J'ai établi les conditions théoriques garantissant la convergence de ces estimateurs. Cependant, ces estimateurs reposant sur des variables aléatoires simulées, leur optimisation est problématique car elle nécessite d'avoir recours aux techniques d'optimisation stochastique très coûteuses en temps de calcul.

Afin de fournir une méthode utilisable en temps réel, j'ai proposé une approche particulière généralisant le principe du filtre avec ré-échantillonnage. L'idée d'utiliser des méthodes de filtrage particulière pour estimer à la fois les variables d'état  $x_t$  non observées du système et les paramètres inconnus constants  $\theta$  du modèle n'est pas nouvelle ([Doucet and Tadić, 2003](#)) pour un état de l'art). Elle repose sur l'introduction d'une loi a priori  $\pi_0(\theta)$  pour les paramètres et une extension du vecteur d'état qui devient  $z_t = (x_t, \theta_t)$ , et consiste à réestimer cette loi par approximation discrète à chaque pas de temps. Elle se heurte généralement au problème d'un modèle d'évolution à introduire pour ces paramètres. Pour les méthodes particulières classiques l'introduction du modèle le plus naturel  $\theta_{t+1} = \theta_t$  restreint l'exploration de l'espace des paramètres à la phase d'initialisation de l'algorithme, empêchant sa convergence. Différentes approches ont été proposées pour pallier cette difficulté. Elles reposent généralement sur l'introduction d'un bruit artificiel dans le modèle des paramètres qui devient  $\theta_{t+1} = \theta_t + \nu_t$ , au risque de dénaturer la dynamique réelle du système.

Pour éviter ces additions une approche par régularisation, de l'estimation discrète de la loi a posteriori des paramètres, a également été proposée ([West, 1993](#)), au prix d'une complexité accrue.

La technique d'estimation de paramètres que j'ai développée à partir du filtre avec ré-échantillonnage peut s'interpréter comme une généralisation de l'approche de [West \(1993\)](#). Comme pour les approches traditionnelles évoquées ci-dessus, le vecteur d'état est étendu au paramètre :  $z_t = (x_t, \theta_t)$ . Cependant, le fait de travailler avec un modèle d'état régularisé permet de respecter formellement le modèle naturel  $\theta_{t+1} = \theta_t$ , pour les paramètres inconnus, sans rencontrer les problèmes habituels de dégénérescence, tout en conservant des propriétés de convergence.

La théorie de cette méthode permettant l'estimation conjointe en temps réel état-paramètre est présentée dans [Rossi and Vila \(2005a\)](#). Lors de mon post-doctorat à l'Inria, j'ai approfondi ce travail pour développer un algorithme utilisable en pratique. L'algorithme SMC que nous avons développé pour estimer des paramètres constants dans des systèmes dynamiques est publié dans l'article [Campillo and Rossi \(2009\)](#).

## 2.2 Inférence de modèles statistiques

L'approche classique (ou fréquentiste) pour estimer les paramètres inconnus d'un modèle est de chercher à maximiser la vraisemblance. Dans le cas de modèles statistiques simples cette approche est très efficace. Mais lorsque les modèles sont plus complexes, il est souvent intéressant des les inférer dans le cadre bayésien car il est toujours possible de construire un algorithme pour inférer le modèle. Ce sont les algorithmes de Monte Carlo par chaîne de Markov (MCMC). Ces algorithmes sont souvent utilisés en écologie ([Brooks et al., 2004](#)), en particulier, nous l'avons proposé pour des modèles de capture-recapture ([Gimenez et al., 2007](#)).

Ces méthodes (MCMC) sont à présent aussi largement utilisées pour étudier l'évolution de ressources naturelles, telles que les biomasses dans les pêcheries ou la dynamiques des forêts. Bien que flexibles ces méthodes convergent très lentement. Les méthodes de filtrage particulière, approches bayésiennes récursives, sont par contre très rapides mais mal adaptées à ce type de modèle où les données sont relevées quotidiennement, mensuellement voir annuellement. Ainsi il y a un réel besoin de proposer un traitement spécifique pour ce type de mesures. Lors de mon post-doc à l'Inria nous avons développé une nouvelle méthode, dans l'esprit des techniques des populations de Monte Carlo récemment introduites, faisant inter-agir des chaînes de Markov semble appropriée. Ces travaux sont détaillés dans la section suivante.

### 2.2.1 Algorithme MCMC en parallèle et en interaction

Les algorithmes de Monte Carlo par Chaîne de Markov (MCMC) (cf. [Robert \(1996\)](#); [Marin and Robert \(2007\)](#)) permettent d'échantillonner suivant une loi de probabilité  $\pi(x) dx$  connue à une constante multiplicative près. Ils consistent à simuler une chaîne de Markov dont la distribution limite est  $\pi(x) dx$ . Par exemple, l'algorithme de Metropolis-Hastings (MH) permet de construire un noyau de transition  $P(x, d\tilde{x})$  réversible et régulier sur  $(\mathbf{E}, \mathcal{E})$  dont  $\pi$  est la mesure invariante. L'algorithme MH fait appel à deux ingrédients :

1. un noyau de transition  $Q(x, d\tilde{x})$  sur  $(\mathbf{E}, \mathcal{E})$ ,
2. une fonction mesurable  $\alpha : \mathbf{E} \times \mathbf{E} \mapsto [0, 1]$ .

Partant d'un état  $x \in \mathbf{E}$ , le noyau  $Q(x, d\tilde{x})$  permet de générer un candidat  $\tilde{x} \in \mathbf{E}$ . Ce candidat sera accepté avec probabilité  $\alpha(x, \tilde{x})$ , dans le cas contraire on conservera l'état  $x$ .

En fait, seul le noyau de proposition  $Q$  doit être donné, l'algorithme MH propose une expression de la probabilité d'acceptation  $\alpha$  de telle sorte que la chaîne de Markov induite admette  $\pi$  comme mesure invariante. Un bon choix de la loi de proposition est essentiel pour obtenir une convergence rapide de l'algorithme. Il existe plusieurs techniques pour accélérer la convergence vers la loi cible en améliorant les propriétés de mélangeance de la chaîne. De plus, le diagnostic de convergence de la méthode est très délicat.

Une alternative est de simuler plusieurs chaînes de Markov en parallèle, la plus simple étant d'utiliser des chaînes parallèles indépendantes. Les avis dans la littérature sur cette alternative sont contradictoires. D'après [Robert \(1996\)](#), des chaînes parallèles indépendantes est une mauvaise idée : parmi toutes les chaînes certaines peuvent ne pas avoir convergées, au quel cas l'estimation obtenue est ainsi dégradée. De plus, des chaînes parallèles indépendantes peuvent exhiber artificiellement un comportement robuste qui ne correspond pas à une convergence réelle de l'algorithme.

Cependant, l'utilisation de chaînes en parallèle qui échangent de l'information pour améliorer les propriétés de mélangeance des algorithmes a fait l'objet de nombreux travaux. Un cadre de travail général «Population Monte Carlo» a été proposé pour décrire ce contexte ([Mengersen and Robert, 2003](#); [Cappé et al., 2004](#)). Nous avons proposé un principe d'interaction entre chaînes en parallèle qui fournit un échantillon indépendant de la loi cible. L'idée est que chaque chaîne propose un candidat pour elle-même mais également pour toutes les autres chaînes.

On montre que l'ensemble de ces  $N$  chaînes en interaction est lui-même une méthode MCMC pour le produit de  $N$  mesures cibles.

Contrairement aux travaux précédemment cités, la loi de proposition est donnée et nous ne cherchons pas à l'adapter durant la procédure. Notre travail ne règle donc pas la question du choix de la loi de proposition. Mais nous avons établi les propriétés théoriques garantissant la convergence des algorithmes intégrant notre principe d'interaction. La question théorique de l'accélération de la convergence reste ouverte, cependant sur deux exemples numériques simples, dont un cas multimodal, nous avons observé que les interactions accélèrent sensiblement la convergence.

Autrement dit, nous avons proposé un nouvel algorithme d'estimation combinant les méthodologies des SMC et des MCMC, permettant de faire interagir des chaînes en parallèle tout en conservant les propriétés de convergence. Ces travaux ont été présentés en conférences internationale et une partie a donné lieu à l'article (Campillo et al., 2009).

### Illustration sur une problématique de pêche

Par exemple, sur une problématique de pêche consistant à estimer le stock de poissons, nous pouvons modéliser l'évolution de la biomasse,  $B_t$ , par un modèle de Ricker :

$$B_{t+1} = B_t \exp(\tau) \exp(-b B_t) \exp(W_t) \quad (2.3)$$

avec  $W_t \sim \mathcal{N}(0, \sigma_w^2)$ . Ce qui donne en passant au log,  $X_t = \log B_t$  :

$$X_{t+1} = X_t + \tau - b \exp(X_t) + W_t \quad (2.4)$$

Le stock de poisson n'est évidemment pas observé directement, mais se déduit à partir de l'effort de pêche. Ce qui est formalisé par un modèle d'observation :

$$Y_t = X_t + V_t \quad (2.5)$$

avec  $V_t \sim \mathcal{N}(0, \sigma_v^2)$ . Nous avons simulé une trajectoire de biomasse sur 20 ans selon le modèle (2.4) et sa série d'observations associées selon le modèle (2.5) en utilisant les paramètres  $\tau = 1.5$ ,  $b = 0.02$ ,  $\sigma_w^1 = 1$ ,  $\sigma_v^2 = 0.5$  et comme état initial du log de la biomasse  $X_1 \sim \mathcal{N}(3, 1)$ . A noter qu'une série de 20 observations est tout à fait conforme au volume de données à disposition pour ce type de problématique. Nous nous sommes mis dans la configuration simpliste où seulement  $\tau$  est à estimer avec pour loi *a priori*  $\pi_\tau^0 = \mathcal{N}(4, 2^2)$ .

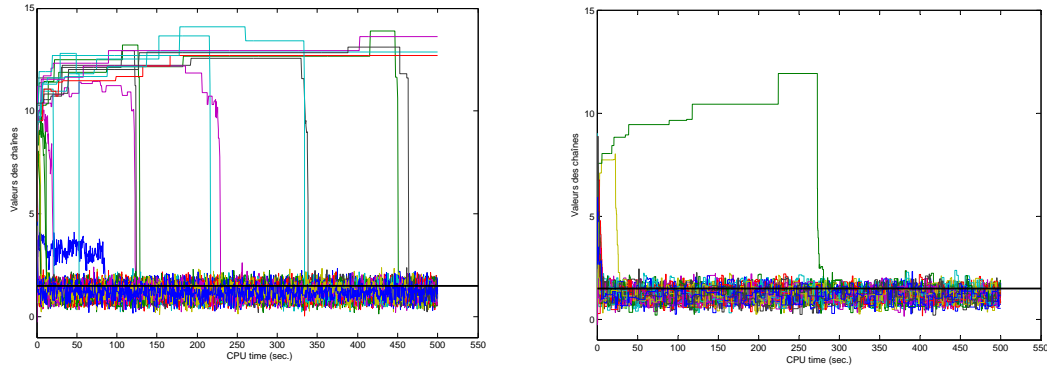


FIGURE 2.1 – Gauche : chaînes MCMC en parallèles pour estimer  $\tau = 1.5$  (ligne horizontale noire); Droite : Chaînes MCMC en parallèle et en interaction pour estimer  $\tau = 1.5$  (ligne horizontale noire)

Comme on peut le voir sur la figure (2.1), les chaînes en interactions convergent beaucoup plus rapidement vers le bassin d’attraction de la vraie valeur de  $\tau$ . L’abscisse des graphiques représente le temps de calcul pour que la comparaison soit équitable. Car une itération de l’algorithme MCMC en parallèle et en interaction est plus coûteuse en temps de calcul qu’une itération de l’algorithme MCMC en parallèle sans interaction. Si, nous comparons les performances des deux algorithmes par rapport au nombre d’itérations, la supériorité de l’algorithme MCMC en parallèle et en interaction est encore plus marquée.

## 2.3 Sélection de modèles multi-réponses

J’ai abordé la question de la sélection de modèle dans le cadre bayésien en parallèle de ma thèse et durant quelques années après. Hors du cas des modèles paramétriques emboîtés, les tests fréquentistes usuels ne sont pas utilisables pour effectuer une sélection entre les modèles. Seuls demeurent les critères de la famille AIC, BIC ou la validation croisée (CV). Cependant, il est établi que ces approches sont inconsistantes (Stone, 1974). La mise en place d’un critère de sélection pour des modèles paramétriques non emboîtés demeure donc une question de recherche très pertinente.

Le problème de la sélection de modèle se formalise généralement de la façon suivante : soit un ensemble de couples  $Z_n = \{(x_i, y_i) \in \mathbb{R}^s \times \mathbb{R}^d; i = 1, \dots, n\}$ , on

cherche la fonction de lien  $f$  telle que  $y_i = f(x_i) + \varepsilon_i$  avec  $\varepsilon_i$   $i = 1, \dots, n$  des bruits i.i.d. centrés de matrice de covariance  $\Sigma$  inconnue. Soit  $\mathcal{M} = \{M^1, \dots, M^m\}$  les différents modèles candidats pour  $f$ . L'objectif de notre problématique est d'établir une hiérarchie entre les différents modèles au regard des données  $Z_n$ . Plus particulièrement, on se restreint au cas où les modèles candidats sont paramétriques :  $\mathcal{M} = \{f_1(x, \theta), f_2(x, \theta), \dots, f_m(x, \theta)\}$ .

Le critère sur lequel j'ai travaillé est basé sur une approche bayésienne où le modèle choisi est celui qui maximise l'espérance d'une fonction d'utilité ([Bernardo and Smith, 1994](#)). En effet, pour la sélection de modèles, il est préférable de prendre une utilité qui est fonction de la densité prédictive des réponses du modèle candidat. L'avantage de ce critère bayésien sur la CV et les variantes d'AIC est sa consistance. Cependant, la forme analytique de cette densité prédictive des réponses est généralement inaccessible, il est donc impossible de calculer le critère d'utilité de manière exacte. Il est alors nécessaire d'approximer la densité prédictive des réponses afin d'obtenir une estimation du critère. J'ai proposé une estimation, asymptotiquement convergente en variation totale, de la densité prédictive des réponses, par une loi de Student en utilisant une loi a priori conjuguée pour les paramètres. Le choix de la loi a priori conjuguée se justifie essentiellement pour contourner des difficultés techniques. Ce résultat appliqué à la sélection de l'architecture d'un réseau de neurones, a donné lieu à des communications dans des congrès et à l'article [Rossi and Vila \(2006a\)](#).

Lors de travaux ultérieurs à ma thèse, j'ai affaibli certaines hypothèses garantissant la consistance du critère. J'ai également étudié les possibilités de l'estimer de manière totalement numérique en utilisant les techniques MCMC : l'algorithme de Metropolis-Hastings et une version hybride de Gibbs et Hastings. L'algorithme de Gibbs classique n'est pas utilisable car aucune contrainte n'est imposée aux modèles candidats. Ces derniers résultats sont publiés dans l'article [Rossi and Vila \(2008\)](#).

## 2.4 Conclusion

En résumé, mes travaux méthodologiques concernent principalement le développement d'algorithmes pour estimer des paramètres inconnus ou des variables non observées dans des modèles dynamiques ou dans des modèles statistiques. J'ai également développé des algorithmes pour estimer un critère de sélection de modèle. Le point commun de tous ces algorithmes est qu'ils sont développés dans

le cadre bayésien, sur des principes soit du type MCMC soit du type SMC. Pour chacun de ces algorithmes, je les ai implémentés et je me suis assuré de leur consistance. C'est à dire que théoriquement, sous des hypothèses usuelles sur les modèles, ils convergent bien vers la quantité que l'on cherche à estimer. Mes projets de nouveaux algorithmes d'inférence sont présentés dans la partie perspective. Ils concernent, plus spécifiquement, des modèles intéressants pour étudier les écosystèmes.

Le chapitre suivant est une synthèse des modèles que j'ai contribué à développer autour des forêts tropicales. Les algorithmes utilisés pour inférer ces modèles sont généralement de type MCMC. Nous n'avons pas utilisé les algorithmes SMC que j'ai développé car ils n'étaient pas les plus pertinents pour nos problématiques. Nous aurions pu utiliser les algorithmes MCMC en parallèle et en interaction. Mais comme les algorithmes MCMC usuels fonctionnaient correctement, je n'ai pas jugé utile d'y avoir recours étant donné leur complexité de mise œuvre. Même si mes algorithmes n'ont pas été utilisés, mes connaissances en la matière ont permis de concevoir des méthodes d'inférence bien adaptées à nos modèles. De plus, avoir travaillé sur des modèles et des données écologiques me donne à présent une vision plus claire des besoins en développements méthodologiques dans ce domaine. Un des nouveaux algorithmes en perspective intègre cette dimension.

### **3 Comment modéliser les forêts tropicales en intégrant leur diversité spécifique ?**

La modélisation se justifie généralement par le besoin d'informations non accessibles directement sur un objet d'étude. Il est possible de classer ces informations en deux catégories, celles décrivant la nature de l'objet et celles décrivant son fonctionnement. Le modèle est une simplification appréhendable de l'objet d'étude qui doit permettre de récupérer les informations désirées.

Dans le cas des forêts tropicales, les informations concernant leurs natures et leurs fonctionnements peuvent être très diverses. Souvent les modèles concernent les services ou la dynamique de cet écosystème. Les services et produits des écosystèmes forestiers sont multiples ([Barnes et al., 1998](#)), par exemple, fourniture en bois d'œuvre, en bois énergie, en produits forestiers non ligneux et à une échelle globale stockage du carbone, réservoir de biodiversité, préservation des fonctions du sol ou stabilité du climat. La connaissance de la dynamique forestière permet quant à elle de prédire comment vont évoluer ces services après une perturbation de la forêt. Les perturbations pouvant être également assez variées ([Barnes et al., 1998](#)), par exemple, exploitation forestière, feu de forêt ou évènement climatique extrême.

Pour mes travaux de modélisation, je me focalise exclusivement sur les arbres de la forêt. C'est à dire les végétaux ligneux dépassant 10 centimètres de diamètre à 1,3 mètres de hauteur. Les modèles n'intègrent donc pas les autres végétaux, les animaux ou les insectes.

Les services de l'écosystème forestier sur lesquels j'ai principalement travaillé sont la biomasse aérienne et le volume en bois d'œuvre. Pour le fonctionnement de la forêt, j'ai travaillé sur des modèles de chacun des processus de la dynamique forestière, c'est à dire la régénération, la croissance et la mortalité des arbres. J'ai



également travaillé sur des modèles de dynamique forestière complets intégrant ces trois processus.

Les questions de recherches abordées lorsque l'on fait de la modélisation sur les forêts tropicales sont en fait dues à leur nature. La principale caractéristique des forêts tropicales est leur grande diversité biologique (Myers et al., 2000). Par exemple, on peut trouver un peu plus de deux cent espèces d'arbre à l'hectare dans les forêts de Guyane française (Marcon et al., 2012). Il n'est donc pas raisonnable de chercher à transposer l'approche des forêts tempérées qui consiste à faire un modèle par espèce. La plupart des espèces sont peu fréquentes, il est donc impossible, hormis pour quelques essences commerciales, d'assembler suffisamment de données pour faire des modèles fiables pour chaque espèce. La prise en compte de cette grande diversité spécifique a constitué le cœur de mes travaux en modélisation sur les forêts tropicales.

La grande diversité spécifique des forêts tropicales est une source de variabilité dans les caractéristiques des arbres et dans leur dynamique (Phillips et al., 1994). Par ailleurs, il existe souvent des données de natures et d'origines diverses contenant des informations utiles pour appréhender cette variabilité. Afin de prendre en compte proprement cette variabilité et ces diverses sources d'information, j'ai construit des modèles dans le cadre bayésien hiérarchique. En effet, la statistique bayésienne est une théorie mathématique bien adaptée à la propagation d'incertitude et l'intégration de données de natures différentes, par exemple elle autorise la prise en compte du savoir subjectif d'experts (Campillo et al., 2008b,a). Un avantage important de la statistique bayésienne est qu'elle permet toujours de construire un algorithme pour calibrer le modèle créée aussi compliqué soit-il. De mon point de vue, un modèle doit être constitué d'un équation définissant la vraisemblance des données, au sens statistique, et d'un algorithme d'inférence adapté aux données à disposition.

### **3.1 Évaluation de la biomasse forestière**

Parmi les services des écosystèmes forestiers tropicaux le stockage de carbone fait parti des services les plus étudiés ces dernières années. En effet la capacité des forêts tropicales humides à stocker le carbone en fait un outil de régulation des émissions de  $CO_2$ . Les risques environnementaux liés aux changements globaux font ainsi des forêts tropicales humides un des acteurs essentiel dans la stratégie mondiale d'atténuation du réchauffement climatique. Dans le cadre des mécanismes

REDD et REDD+, la valeur potentielle des forêts vue comme un stock ou un puits de carbone est généralement supérieure à celle des forêts vue comme fournisseurs de biens et services ([Ebeling and Yasué, 2008](#)).

L'évaluation du stock de carbone ou de la biomasse des forêts tropicales est donc un enjeu majeur. Une estimation fiable des stocks de carbone aérien dans les écosystèmes forestiers repose d'une part sur un large jeu de données (couvertures forestières, mesures dendrométriques, identification des arbres, connaissance de la densité de bois, etc.) ; d'autre part sur un modèle de propagation des incertitudes lorsque l'on passe de la biomasse d'un arbre, à celle d'une parcelle puis lorsque l'on extrapole à une région forestière de plusieurs dizaine de kilomètres carrées.

Ces données, si elles sont disponibles sur les écosystèmes forestiers tempérés, héritage d'une longue tradition de recherche forestière, sont le plus souvent rares en zones tropicales. La mise au point de modèles est également plus difficile pour les écosystèmes forestiers tropicaux en raison du nombre très élevé d'espèces d'arbres.

Aujourd'hui une grande attention est portée sur les méthodes basées sur des données de télédétection ([Goetz et al., 2009](#)). Ces approches donnent déjà de bons résultats sur les forêts tempérées mais rencontrent encore quelques difficultés sur les forêts tropicales car leurs stocks de carbone sont très élevés ce qui cause la saturation du signal. Lorsqu'elles seront opérationnelles, elles permettront d'avoir quasi-automatiquement des estimations de stock de carbone à l'échelle d'un pays à des pas de temps réguliers. Cependant, elles nécessiteront toujours d'avoir des points de référence au sol, où le stock de carbone est connu, pour calibrer ou recalibrer les capteurs des satellites.

Or l'incertitude des estimations de stockage de carbone, obtenues à partir des données de dispositifs forestiers permanents de quelques hectares, étaient rarement précisées ([Chave et al., 2004](#)). Alors que pour raisonner sur les variations de stock de carbone et donc sur le rôle de puits ou de source de carbone des forêts, il est nécessaire d'avoir un intervalle de confiance pour chacune des estimations.

Mes travaux sur l'estimation du stock de carbone ont principalement consisté à construire rigoureusement des intervalles de confiance pour les estimations de stock de carbone. Pour cela, il est essentiel de prendre en compte la qualité et la quantité de données disponibles pour quantifier l'incertitude induite par ces données. Pour être précis, j'ai travaillé sur l'estimation de la biomasse aérienne sèche de la forêt. Il est communément admis que le stock de carbone équivaut à

environ 50% de la biomasse sèche (Martin and Thomas, 2011).

### 3.1.1 Estimation de la biomasse d'une parcelle à partir d'inventaires d'exploitants forestiers

Les premiers modèles que j'ai construits pour estimer la biomasse forestière étaient destinés à étudier les variations de la biomasse suivant le type de sol dans les forêts de République centrafricaine (Gourlet-Fleury et al., 2011). Nous avons les données d'inventaires des exploitants forestiers : les effectifs d'arbres par classes de diamètres au delà de 20 cm sur 2754 placettes de 0.5 hectares répartis sur 4888  $km^2$ . Il y avait au total 209 espèces d'arbre inventoriées sur toutes les parcelles.

#### Modèle de biomasse pour un arbre

Pour estimer la biomasse  $Ba$  d'un arbre, nous avons choisi d'utiliser l'allométrie de Chave et al. (2004) :

$$\begin{aligned} Ba(\theta, wsg, d) &= wsg \exp\{\theta_0 + \theta_1 \log dbh + \theta_2 (\log dbh)^2 + \theta_3 (\log dbh)^3\} \varepsilon \\ &= wsg g(\theta, dbh) \varepsilon \end{aligned}$$

avec  $dbh$  le diamètre de l'arbre à 1,30 m ,  $wsg$  la densité du bois de l'espèce à laquelle appartient l'arbre et  $\log \varepsilon \sim \mathcal{N}(0, \sigma^2)$ . Nous avons supposé que les paramètres  $\theta$  et  $\sigma$  obtenus par Chave et al. (2004) étaient corrects. Cependant, le diamètre était incertain puisque nous connaissions uniquement la classe diamétrique de l'arbre. La densité du bois  $wsg$  était également inconnue. Nous avons construit, en utilisant la base de données des densités des bois du Cirad et la phylogénie, un modèle de connaissance afin d'avoir l'espérance,  $\mathbb{E}[wsg]$ , et la variance,  $V[wsg]$ , de la densité du bois pour chaque espèce. Il est alors possible pour toutes les espèces de calculer l'espérance et la variance de la biomasse d'un de ces arbres sachant son diamètre  $dbh$ . Pour cela il suffit d'utiliser les propriétés de l'espérance conditionnelle suivantes :

$$\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y] \quad \text{et} \quad V[Y] = \mathbb{E}[V[Y|X]] + V[\mathbb{E}[Y|X]].$$

Ce qui appliqué au modèle de biomasse donne pour l'espérance :

$$\begin{aligned} \mathbb{E}[Ba|dbh] &= \mathbb{E}[wsg g(\theta, dbh) \varepsilon] \\ &= g(\theta, dbh) \mathbb{E}[wsg] \mathbb{E}[\varepsilon] \end{aligned}$$

Pour la variance

$$\begin{aligned}
V[Ba|dbh] &= \mathbb{E}_{wsg} [V[wsg g(\theta, dbh) \varepsilon | wsg, dbh] | d] \\
&\quad + V_{wsg} [\mathbb{E}[wsg g(\theta, dbh) \varepsilon | wsg, dbh] | dbh] \\
&= \mathbb{E}_{wsg} [wsg^2 g(\theta, dbh)^2 V[\varepsilon] | dbh] + V_{wsg} [wsg g(\theta, dbh) \mathbb{E}[\varepsilon]] \\
&= \mathbb{E}[wsg^2] g(\theta, dbh)^2 V[\varepsilon] + V[wsg] g(\theta, dbh)^2 \mathbb{E}[\varepsilon]^2 \\
&= [\mathbb{E}[wsg^2] V[\varepsilon] + V[wsg] \mathbb{E}[\varepsilon]] g(\theta, dbh)^2
\end{aligned}$$

Il reste encore à intégrer l'incertitude sur le diamètre. Pour chaque parcelle, on a la distribution diamétrique des arbres pour les espèces présentes. Soit  $L$  le nombre de classes diamétriques. Pour la parcelle  $i$ , où il y a  $S_i$  espèces présentes, les données sont  $n_{l,s}$  l'effectif de l'espèce  $s$  dans la classe diamétrique  $l$  pour  $s = 1, \dots, S_i$  et  $l = 1, \dots, L$ . Pour déterminer numériquement l'espérance et la variance de la biomasse d'un arbre de n'importe quelle espèce dans n'importe quelle classe de diamètres, il suffit de connaître les valeurs numériques de

$$\mathbb{E}_{dbh \in l} [g(\theta, dbh)] \quad \text{et} \quad \mathbb{E}_{dbh \in l} [g(\theta, dbh)^2].$$

Pour fixer les idées, si la distribution diamétrique est exponentielle de paramètre  $\alpha$  :

$$\mathbb{E}_{dbh \in l} [g(\theta, dbh)] = \frac{\alpha \int_{bin f_l}^{bsup_l} g(\theta, dbh) \exp(-\alpha dbh) ddbh}{\exp(-\alpha bin f_l) - \exp(-\alpha bsup_l)}$$

et

$$\mathbb{E}_{dbh \in l} [g(\theta, dbh)^2] = \frac{\alpha \int_{bin f_l}^{bsup_l} g(\theta, dbh)^2 \exp(-\alpha dbh) ddbh}{\exp(-\alpha bin f_l) - \exp(-\alpha bsup_l)}$$

avec respectivement  $bin f_l$ ,  $bsup_l$  les limites inférieure et supérieure de la classe diamétrique  $l$ .

Il est possible de préciser analytiquement les deux premiers moments de la biomasse d'un arbre de n'importe quelle espèce sachant sa classe de diamètre  $l$  :

$$\begin{aligned}
\mathbb{E}[Ba|dbh \in l] &= \mathbb{E}_{dbh \in l} [\mathbb{E}[Ba|dbh]] \\
&= \mathbb{E}_{dbh \in l} [g(\theta, dbh) \mathbb{E}[wsg] \mathbb{E}[\varepsilon]] \\
&= \mathbb{E}_{dbh \in l} [g(\theta, dbh)] \mathbb{E}[wsg] \mathbb{E}[\varepsilon]
\end{aligned}$$

On en déduit également la variance

$$\begin{aligned}
V[Ba|dbh \in l] &= \mathbb{E}_{dbh \in l} [V[Ba|dbh]] + V[\mathbb{E}[Ba|dbh]] \\
&= \mathbb{E}_{dbh \in l} [\mathbb{E}[wsg^2] V[\varepsilon] + V[wsg] \mathbb{E}[\varepsilon]] g(\theta, dbh)^2 \\
&\quad + V[g(\theta, dbh) \mathbb{E}[wsg] \mathbb{E}[\varepsilon]] \\
&= \mathbb{E}_{dbh \in l} [g(\theta, dbh)^2] [\mathbb{E}[wsg^2] V[\varepsilon] + V[wsg] \mathbb{E}[\varepsilon]] \\
&\quad + \mathbb{E}[wsg]^2 \mathbb{E}[\varepsilon]^2 V[g(\theta, dbh)]
\end{aligned}$$

### Modèle de biomasse pour une parcelle

Pour passer à la biomasse d'une parcelle, il suffit simplement de sommer les biomasses des arbres de la parcelle. Par conséquence, la loi de probabilité de la biomasse dans la parcelle  $i$ ,  $Bp_i$ , sachant les  $n_{l,s}$  est le produit de convolution des  $\pi_{Ba|l,s}$  dilatées par leurs effectifs associés  $n_{l,s}$ . Dans ce cas on pourra toujours récupérer l'espérance et la variance :

$$\mu_i = \mathbb{E}[Bp_i] = \sum_{s=1}^{S^i} \sum_{l=1}^L n_{l,s} \times \mathbb{E}[Ba|dbh \in l, s]$$

$$\sigma_i^2 = V[Bp_i] = \sum_{s=1}^{S^i} \sum_{l=1}^L n_{l,s} \times V[Ba|dbh \in l, s]$$

Si on considère qu'il y a de l'incertitude sur les effectifs observés  $n_{l,s}$ , notons  $N_{l,s}$  les effectifs théoriques, alors l'espérance et la variance de la biomasse dans la parcelle  $i$ ,  $Bp_i$ , sachant les  $n_{l,s}$  sont définies par

$$\mu_i = \mathbb{E}[Bp_i] = \sum_{s=1}^{S^i} \sum_{l=1}^L \mathbb{E}[N_{l,s}|n_{l,s}] \mathbb{E}[Ba|dbh \in l, s]$$

$$\sigma_i^2 = V[Bp_i] = \sum_{s=1}^{S^i} \sum_{l=1}^L \mathbb{E}[N_{l,s}^2|n_{l,s}] \mathbb{E}[Ba^2|dbh \in l, s] - \mathbb{E}[N_{l,s}|n_{l,s}]^2 \mathbb{E}[Ba|dbh \in l, s]^2$$

Le cadre formel, que j'ai défini ci-dessus, m'a permis d'estimer proprement l'espérance et la variance de la biomasse des parcelles. Ainsi nous avons pu mettre en évidence, en toute rigueur, une différence significative de la biomasse aérienne suivant le type de sol. La propagation des incertitudes se fait ici exclusivement par l'intermédiaire de la variance. Il aurait évidemment été plus intéressant d'avoir la distribution de la biomasse plutôt que l'espérance et la variance. Comme nous n'avons pas les lois *a posteriori* des paramètres obtenus dans [Chave et al. \(2004\)](#) pour l'allométrie de biomasse et que nous n'avons pas de données de masses d'arbres pour les recalculer ce n'était pas possible.

### 3.1.2 Estimation de la biomasse d'une parcelle à partir d'inventaires scientifiques

Je me suis servi du cadre formel précédent pour définir le sujet la thèse de Quentin Molto que j'ai co-encadré. Cette fois, nous avons pu propager rigoureusement

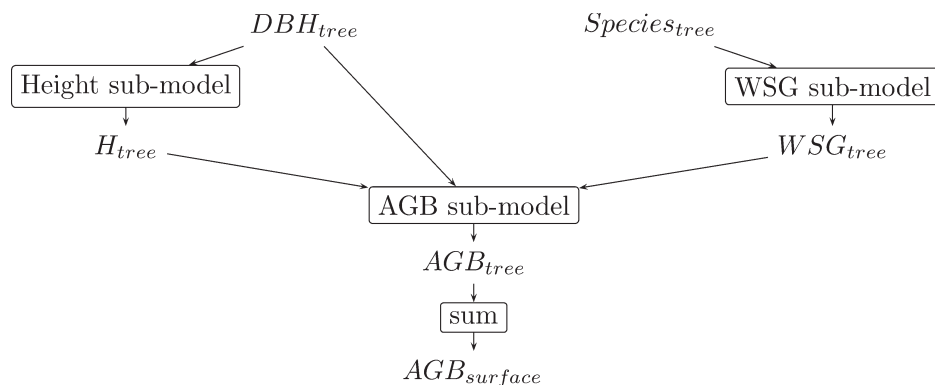


FIGURE 3.1 – Diagramme du modèle de propagation des incertitudes dans l’estimation de la biomasse d’une parcelle (extrait de Molto et al. (2013))

toutes les sources d’incertitude et quantifier leurs importances relatives pour l’estimation de la biomasse d’une parcelle.

Cela a été réalisable car, en Guyane française, de nombreux jeux de données sont disponibles. Ces données varient selon l’échelle spatiale : du local (dispositifs forestiers de quelques hectares) au régional (inventaire forestier couvrant des milliers d’hectares). Ainsi sur quelques hectares de forêts, nous possédons un jeu de données extrêmement précises issu du projet Bridge (identification à l’espèce de tous les arbres, mesure de sa densité du bois, mesure de sa circonférence et de sa hauteur, <http://www.ecofog.gf/Bridge>). A l’extrême, les inventaires forestiers des années 60 et 70 couvrent de très grandes surfaces mais les informations sur les arbres sont peu précises (mesure à l’œil de la circonférence, nom vernaculaire pour les essences commerciales, . . .). Mais, le plus important est que nous avons pu accéder à des données d’arbres pesés (Lescure et al., 1983).

Nous avons conservé le même schéma que dans les cas de données d’inventaires d’exploitants forestiers. A savoir, nous avons construit un modèle de biomasse à l’échelle de l’arbre, puis nous sommes passés à la biomasse de la parcelle en agrégeant les prédictions de biomasse des arbres. Mais comme les données à disposition nous permettaient de le faire, nous avons construit un modèle bayésien hiérarchique pour intégrer les différentes sources d’erreurs. Nous avons inféré le modèle de biomasse d’un arbre récupérant ainsi les lois a posteriori de ces paramètres. Nous avons ainsi pu construire la distribution de probabilité de la biomasse

d'une parcelle en prenant en compte toutes les incertitudes. Ce travail a constitué une première en milieu tropical, il est présenté dans l'article [Molto et al. \(2013\)](#).

### Modèle de biomasse pour un arbre

Les modèles de biomasse des arbres sont généralement des équations proches de la masse d'un cylindre. C'est à dire qu'elle font intervenir la hauteur,  $h$ , le diamètre,  $dbh$ , et la densité de du bois de l'arbre,  $wsg$ . Comme la variance des biomasses augmentent avec le diamètre de l'arbre, les modèles prédisent généralement le log de la biomasse. Nous avons suivi cette logique en modélisant la biomasse de l'arbre  $t$ ,  $Ba_t$  par l'équation :

$$\log Ba_t = \theta_0 + \theta_1 h_t + \theta_2 dbh_t + \theta_3 wsg_t + \varepsilon_t \quad (3.1)$$

avec  $\varepsilon_t$  distribuée selon une loi normale centrée de variance  $\sigma^2$ .

A première vue, ce modèle (3.1) pourrait être inféré facilement avec un algorithme MCMC de type Gibbs. Cependant, toutes les variables explicatives ne sont pas renseignées dans le jeu de données des arbres pesés ([Lescure et al., 1983](#)). En effet, la masse, le diamètre et la hauteur des arbres sont renseignés mais pas la densité du bois. Nous avons donc construit un modèle pour prédire la densité du bois d'un arbre en fonction de son espèce à partir des données récoltées dans le projet Bridge. Ce modèle de densité du bois a été inspiré par le modèle de densité du bois que j'ai construit dans ([Gourlet-Fleury et al., 2011](#)). Mais l'information phylogénique a été propagée plus rigoureusement cette fois : nous l'avons fait avec un modèle bayésien hiérarchique. Ainsi pour chaque espèce d'arbre nous avons obtenu une distribution de probabilité pour la densité de son bois.

Ensuite, nous avons pu inférer le modèle (3.1) en prenant en compte l'incertitude sur la densité du bois. Pratiquement, nous avons couplé un schéma de Monte-Carlo à l'algorithme de Gibbs. C'est à dire à chaque itération de l'algorithme de Gibbs nous ré-échantillonnions les valeurs des densités du bois des arbres leurs distributions de probabilité. Nous avons ainsi obtenu la loi a posteriori pour tous les paramètres du modèle :  $\theta_0, \theta_1, \theta_2, \theta_3$  et  $\sigma^2$ .

Nous étions ainsi en mesure d'estimer la distribution de probabilité de la biomasse d'un arbre dont nous connaissions, l'espèce, le diamètre et la hauteur. Or la hauteur n'est généralement pas mesurée dans les inventaires forestiers. Nous avons alors construit une allométrie diamètre-hauteur pour être capable de prédire la hauteur des arbres à partir de leur diamètre. Cette allométrie a été modélisée dans

un cadre bayésien pour obtenir une distribution de probabilité pour la hauteur d'un arbre connaissant son diamètre.

Ainsi pour tout arbre dont nous connaissions au moins le diamètre, en utilisant le modèle (3.1) et des schémas de Monte Carlo pour les quantités incertaines, nous sommes en mesure de produire une estimation de la distribution de probabilité de sa biomasse. Ces estimations des distributions de probabilité de biomasse correspondent en pratique à des échantillons de valeurs issues des distributions de probabilité de biomasse. Ainsi pour un arbre  $t$  nous avons un ensemble de valeurs de biomasse  $Ba_t^1, \dots, Ba_t^N$  avec  $N$  aussi grand qu'on le souhaite.

### Modèle de biomasse pour une parcelle

Soit une parcelle forestière constituée de  $T$  arbres. Le modèle de biomasse pour un arbre défini dans la section précédente, nous permet pour chacun de ces arbres d'obtenir une estimation de la distribution de probabilité de leurs biomasses définie par un ensemble de valeurs de biomasse. Pour estimer la distribution de probabilité de la biomasse d'une parcelle nous générons aussi un échantillon de valeurs issu de cette distribution. Pour générer une valeur de biomasse de la parcelle issue de sa distribution de probabilité, il ne serait pas correct de simplement tirer aléatoirement pour chaque arbre  $t$  une valeur de biomasse dans son ensemble  $\{Ba_t^1, \dots, Ba_t^N$  et de sommer toutes les valeurs tirées pour chacun des arbres. En effet, il faut prendre en compte la dépendance induite par les paramètres du modèle de biomasse arbre (3.1). L'algorithme d'estimation de la distribution de probabilité de la biomasse d'une parcelle pour se résumer par les étapes suivantes

1. génération d'une valeur pour chacun des paramètres du modèle (3.1) selon leur loi a posteriori :  $(\tilde{\theta}_0, \dots, \tilde{\theta}_3, \tilde{\sigma}) \sim \pi(\theta_0, \dots, \theta_3, \sigma | Data)$
2. pour chaque arbre  $t, t = 1 : T$ , génération d'une valeur de densité du bois selon le modèle de densité du bois :  $\tilde{w}sg_i \sim \pi(wsp | espece\ de\ t)$
3. si nécessaire, pour chaque arbre  $i$ , génération d'une valeur de hauteur selon le modèle diamètre-hauteur :  $\tilde{h}_i \sim \pi(h | dbh)$
4. pour chaque arbre  $t$ , génération d'une valeur de biomasse  $\tilde{B}a_t$  en utilisant les valeurs  $\tilde{\theta}_0, \dots, \tilde{\theta}_3, \tilde{\sigma}, \tilde{h}_i, \tilde{w}sg_i$  dans le modèle (3.1).
5. la réalisation de la biomasse de la parcelle est obtenue en sommant toutes les biomasses des arbres :  $\tilde{B}p = \sum_{t=1}^T \tilde{B}a_t$

En répétant ces étapes autant de fois que désiré, nous obtenons un ensemble de valeurs de biomasse pour la parcelle  $\{\tilde{B}p^1, \dots, \tilde{B}p^N\}$  qui constitue l'estimation



de la distribution de la biomasse de la parcelle. A partir de cet ensemble de valeur, il est facile de récupérer les statistiques usuelles, moyenne, médiane ou intervalles de confiance.

Ce modèle a permis de déterminer le niveau de précision que l'on peut atteindre pour l'estimation de la biomasse suivant la qualité des données dont on dispose. Aussi, il apporte un éclairage sur le degré de significativité des variations de biomasses rapportées dans la littérature ces dernières années (Lewis et al., 2004; Phillips et al., 2004a).

Ce modèle a également permis de hiérarchiser les sources d'incertitudes dans les estimations de biomasse. Il est ressorti que les incertitudes apportées par la densité du bois et par la hauteur sont négligeables par rapport à celle du modèle. Cependant, un biais dans les données de hauteurs, ce qui est assez courant en forêt tropicale, peut induire un biais important dans les estimations de biomasse.

### **Modèle de biomasse pour la forêt guyanaise**

Pour atteindre l'objectif de la thèse de Quentin Molto, il nous fallait encore extrapoler les estimations des parcelles permanentes à toute la forêt guyanaise. Nous avons tout d'abord étudié les liens entre la biomasse des parcelles et des variables environnementales. Les variables environnementales n'ont pas permis de bien expliquer les variations de biomasse observées sur les parcelles. Nous avons cherché des relations non linéaires mais sans succès. Nous avons donc décidé de rester sur un modèle linéaire, qui même s'il présente un pouvoir explicatif faible, avait pour lui l'avantage d'être simple. Nous avons ainsi sélectionné sur un modèle linéaire les variables environnementales à conserver pour faire l'extrapolation spatiale. Il s'agissait : de la géologie (cf. carte de Delor et al. (2001)), du type de végétation (cf. carte de Gond et al. (2011)), de la topographie (pente et rugosité à partir des données STRM de la Nasa), de l'hydrographie (aire drainée et altitude par rapport au cours d'eau le plus proche à partir des données SRTM de la Nasa) et du climat (pluviométrie et indice de sécheresse à partir des données TRMM de la Nasa). Ainsi nous avons utiliser un modèle de régression spatial de type co-krigeage (Diggle and Ribeiro, 2007). Sur chaque carré de forêt sur la carte de la Guyane, la biomasse est prédite par un loi normale dont la moyenne est une combinaison linéaire des variables environnementales et la variance intègre l'auto-corrélation spatiale de la biomasse.

Afin de prendre en compte l'incertitude des estimations de biomasse des parcelles pour construire la carte de la biomasse, nous avons réalisé un schéma de Monte

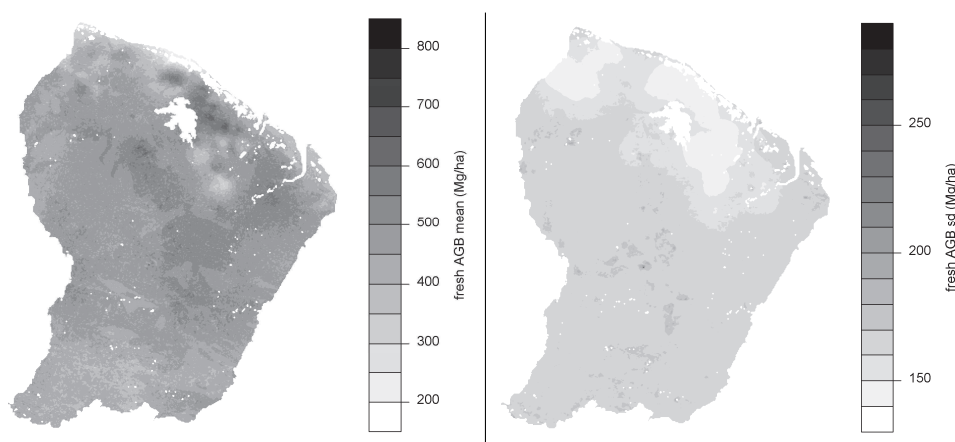


FIGURE 3.2 – Gauche : carte de la biomasse aérienne fraîche de la Guyane ; Droite : carte des écart-types de ces estimations

Carlo. Nous avons échantillonné une valeur de biomasse pour chacune des parcelles suivant leurs loi a posteriori puis nous avons réalisé la régression spatiale afin d’obtenir une estimation de la moyenne et de la variance de la biomasse en tout point de la carte. Ces deux étapes ont été répétées de nombreuses fois obtenir les cartes de la figure (3.2).

Nous avons ainsi pu réaliser la première carte de la biomasse de Guyane par propagation des incertitudes lors de toutes les étapes de changement d’échelle en partant d’un modèle biomasse pour un arbre. Ce travail a nécessité d’intégrer des informations de type très variés dans différents modèles et ensuite de connecter tous ces modèles. Toutes les étapes de ce travail ont été faites dans un cadre cohérent et rigoureux statistiquement grâce à la modélisation bayésienne. Il me semble peu vraisemblable qu’un travail aussi abouti puisse être mené ailleurs qu’en Guyane car cela nécessite d’avoir un volume et une diversité de données peu commun en zone tropicale.

### 3.2 Modélisation de la dynamique des forêts tropicales à l’échelle de la communauté

Les forêts tropicales sont parmi les écosystèmes terrestres les plus diversifiés, en particulier pour la floristique (Swaine and Whitmore, 1988). Il existe différentes théories en écologie pour expliquer de tels niveaux de biodiversité, notamment les deux les plus communément discutées : la théorie neutraliste (Hubbell, 2001) et la

théorie des niches écologiques ([Chase and Leibold, 2003](#)). Tout d'abord, ces deux théories ont été mises en opposition, puis il est à présent supposé que les deux processus écologiques opèrent chacun à des pas de temps et d'espace propres simultanément dans les systèmes biologiques ([Hérault, 2007](#)).

Mon travail n'a pas consisté à chercher une nouvelle théorie expliquant cette biodiversité ou à tester les théories établies mais à étudier le fonctionnement des forêts en intégrant cette biodiversité. L'objectif étant de prédire leurs réactions face à des perturbations variées. Pour y parvenir, la voie privilégiée est de réduire la dimension de cette diversité. J'ai abordé cette question sous deux angles différents : les groupes fonctionnels et les traits fonctionnels. Dans le cas des groupes fonctionnels, les espèces ayant un comportement similaire sont regroupées. La dimension de la diversité est ainsi réduite au nombre de groupes. Dans le cas des traits fonctionnels, les espèces sont projetées dans l'espace engendré par les traits. La dimension de la biodiversité est ainsi réduite au nombre de traits utilisés. Mes travaux relatifs à chacune de ces approches sont présentés dans les deux sections suivantes. Mon analyse des avantages et inconvénients de ces approches fait l'objet de la dernière section de cette partie sur la modélisation de la dynamique forestière.

Avant de rentrer dans le détail de mes travaux, il faut préciser qu'initialement, les perturbations ciblées étaient essentiellement les différentes pratiques d'exploitation forestière. Car étant chercheur au Cirad, la finalité des questions que j'aborde intègre les demandes sociales. Ainsi mes choix initiaux de modélisation ont été faits dans cette optique. La forte montée de la demande sociale sur l'impact des changements climatiques m'a incité à également cibler cet aspect par la suite. Les choix de modélisation en dynamiques forestières concernent l'échelle de modélisation de la forêt. C'est à dire, il faut choisir un niveau d'agrégation des individus entre les deux extrêmes, l'arbre (pas d'agrégation) et le peuplement (agrégation de tous les arbres). Cette question a été abordée en profondeur dans la littérature ([Picard and Franc, 2001](#)). Or pour les questions relatives à l'exploitation forestière, les modèles matriciels (répartition des arbres dans des classes diamétriques) sont efficaces et bien adaptés aux données des inventaires forestiers. Par contre les modèles matriciels usuels ne sont pas les mieux adaptés pour explorer finement les relations entre la dynamique forestière et le climat.

Mes premiers travaux pour appréhender la biodiversité ont concerné l'approche par groupes fonctionnels. Comme ma principale cible était alors de prédire la réaction du peuplement face à l'exploitation, j'ai ainsi travaillé sur des modèles matriciels. Lorsque j'ai commencé à chercher à utiliser les traits fonctionnels pour

appréhender la biodiversité, j'avais déjà en ligne de mire l'intégration du climat dans les modèles de dynamique forestière, ainsi j'ai travaillé sur des modèles à l'échelle de l'arbre.

### 3.2.1 Approche par groupes fonctionnels

L'idée sous-jacente aux groupes fonctionnels est qu'il existe des caractéristiques facilement mesurables, ou des traits, permettant d'identifier le fonctionnement d'une espèce face aux perturbations ([Gitay and Noble, 1997](#); [Lavorel and Garnier, 2002](#)). Si cette théorie est valide, il ne serait donc plus nécessaire d'établir une classification phylogénique complète des espèces d'arbres tropicaux. Cette perspective serait très intéressante en forêt tropicale car la détermination botanique complète de parcelles forestières à un coût très important. Cela supposerait aussi que ces groupes ont une certaine universalité et qu'ils devraient être assez facilement identifiables quel que soit le type de données utilisées. Par induction, toutes les classifications proposées sur une communauté devraient être assez similaires ([Gitay and Noble, 1997](#)).

Outre ces aspects fondamentaux sur l'intérêt et l'existence des groupes fonctionnels, nous sommes aussi soumis à une contrainte opérationnelle forte. Dans les forêts tropicales la plupart des espèces ont peu de représentants, même pour de grands dispositifs de suivi, car la diversité d'espèces à l'hectare est grande. Il est ainsi impossible de calibrer un modèle de dynamique par espèce.

Comme expliqué en préambule, j'ai travaillé sur la construction de groupes fonctionnels sur des modèles matriciels. Les modèles matriciels ([Caswell, 2001](#)) sont des modèles de dynamique des populations qui intègrent les trois processus recrutement, croissance et mortalité. Les individus sont répartis dans des classes ordonnées, les paramètres de ce modèle sont : le taux de recrutement dans la première classes, les taux de passage dans les classes supérieurs et les taux de mortalités dans chaque classe. Dans le cas de la foresterie, ce modèle général se simplifie en n'autorisant la passage d'un individu que dans la classe qui lui est immédiatement supérieure, il s'agit du modèle d'Usher ([Usher, 1969](#)). Cette contrainte offre l'avantage de réduire considérablement le nombre de paramètres à estimer et n'est pas gênante avec les arbres car il est assez simple d'adapter la taille des classes diamétriques pour s'assurer qu'entre deux inventaires aucun arbre ne puisse sauter deux classes.

Dans sa version classique le modèle d'Usher prédit l'espérance des effectifs dans chacune des classes. Afin de prendre en compte la stochasticité démographique et

ainsi pouvoir récupérer proprement des intervalles de confiance, nous avons travaillé avec la version stochastique du modèle d’Usher. Son principe est brièvement rappelé avant d’expliquer comment nous avons construit des groupes fonctionnels à partir de ces modèles d’Usher stochastiques.

### Modèle d’Usher stochastique

La population des arbres est décrite par un vecteur  $\vec{N}_t = (N_{t,1}, \dots, N_{t,L})$  constitué des effectifs dans les  $L$  classes diamétriques ordonnées à l’instant  $t$ . On note  $N_t = \sum_i N_{t,i}$  l’effectif total de la population. Comme tous les modèles de population matriciels, le modèle d’Usher peut s’exprimer comme l’espérance de  $N_t$  chaînes de Markov indépendantes (cf. figure 3.3). Ainsi la relation entre  $\vec{N}_{t-1}$  et  $\vec{N}_t$  est définie par une matrice  $L \times L$ ,  $\mathbf{U}$  appelée matrice d’Usher :

$$\mathbb{E}[\vec{N}_t | \vec{N}_{t-1}] = \mathbf{U} \mathbb{E}[\vec{N}_{t-1}] \quad (3.2)$$

avec

$$\mathbf{U} = \begin{pmatrix} p_1 + f & f & \dots & f \\ q_1 & p_2 & & 0 \\ & \ddots & \ddots & \\ 0 & & q_{L-1} & p_L \end{pmatrix}.$$

$p_l$  est la probabilité pour un arbre de rester dans la classe  $l$ ,  $q_l$  la probabilité de passer de la classe  $l$  à la classe  $l + 1$  et  $f$  la fécondité moyenne.  $q_l$  et  $p_l$  sont à valeurs dans  $[0, 1]$ , alors que  $f$  est à valeurs dans  $\mathbb{R}_+$ . La probabilité de mourir pour un arbre dans la classe  $l$  est donc  $m_l = 1 - p_l - q_l$ . Ainsi la trajectoire arbre est modélisé par une chaîne de Markov 3.3 schématisée par la figure (3.3)

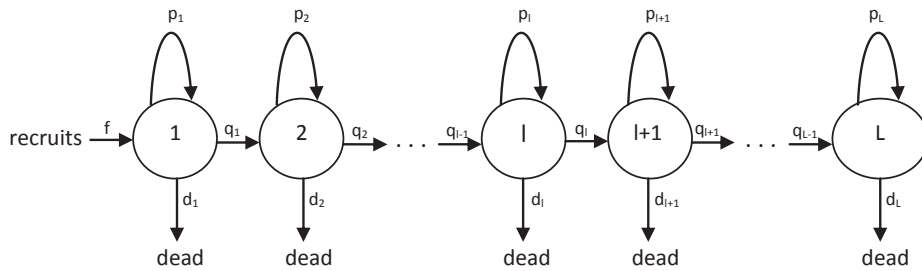


FIGURE 3.3 – Schéma du modèle de Usher stochastique

Comme l'objectif est de regrouper les espèces, il faut aussi intégrer des paramètres relatifs à la distribution diamétrique des espèces,  $\vec{d} = (d_1, \dots, d_L)$  avec  $\sum_{l=1}^L d_l = 1$ . Ainsi  $d_l$  désigne la probabilité qu'un arbre choisi au hasard dans la population soit dans la classe diamétrique  $l$ . Afin de pouvoir écrire la vraisemblance du modèle, il faut considérer les données dans une forme différente intégrant l'évolution des arbres entre les instants  $t - 1$  et  $t$  : soit  $N_{l,l,t}$  le nombre d'arbres qui sont restés dans la classe  $l$  entre  $t - 1$  et  $t$ ,  $N_{l,l+1,t}$  le nombre d'arbres qui sont passés de la classe  $l$  à  $l + 1$  entre  $t - 1$  et  $t$ , et  $N_{l,\dagger,t}$  le nombre d'arbres qui sont morts dans la classe  $l$  entre  $t - 1$  et  $t$ . Soit  $R_t$  le nombre de recrutés en  $t - 1$  et  $t$ , supposé suivre une loi de Poisson de paramètre  $fN_{t-1}$ . L'ensemble de toutes les données entre  $t - 1$  et  $t$  est  $Data_{t|t-1} = \vec{N} = (N_{1,l,t}, \dots, N_{L,\dagger,t}, \vec{N}_{t-1}, R_t)$ . On a ainsi tous les éléments pour écrire la vraisemblance du modèle d'Usher stochastique :

$$\begin{aligned} \mathcal{L}(Data_{t|t-1}|\theta) &= \prod_{l=1}^{L-1} \mathcal{M}(N_{l,l,t}, N_{l,l+1,t}, N_{l,\dagger,t} | p_l, q_l, m_l, N_{l,t-1}) \quad (3.3) \\ &\quad \times \mathcal{M}(N_{L,L,t}, N_{L,\dagger,t} | p_L, m_L, N_{L,t-1}) \\ &\quad \times \mathcal{M}(N_{1,t-1}, \dots, N_{L,t-1} | d_1, \dots, d_L, N_{t-1}) \\ &\quad \times \mathcal{P}(R_t | fN_{t-1}) \end{aligned}$$

où  $\mathcal{M}$  désigne la densité de la loi multinomiale,  $\mathcal{P}$  la densité de la loi de Poisson et  $\theta = (\vec{p}, \vec{q}, \vec{m}, f, \vec{d})$  le vecteur de tous les paramètres inconnus :  $\vec{p} = (p_1, \dots, p_L)$ ,  $\vec{q} = (q_1, \dots, q_{L-1})$  and  $\vec{m} = (m_1, \dots, m_L)$ .

Pour inférer le modèle d'Usher stochastique, nous nous sommes placés dans le cadre bayésien. D'après l'équation de la vraisemblance (3.3), on peut inférer séparément le processus de recrutement. Ainsi en utilisant la loi Gamma  $\mathcal{G}(\alpha_0, \beta_0)$  comme loi *a priori* pour le paramètre  $f$ , nous sommes dans le cas des lois conjuguées et nous pouvons ainsi récupérer directement la loi *a posteriori* pour le paramètre  $f$ . Les dépendances entre paramètres et données pour le processus de recrutement sont représentées sur la figure (3.4).

Pour les processus de croissance et mortalité, en utilisant une loi de Dirichlet  $\mathcal{D}(\delta_p, \delta_q, \delta_d)$  comme loi *a priori* pour les paramètres  $(p_l, q_l, m_l)$ , nous sommes dans le cas des lois conjuguées et nous pouvons ainsi récupérer directement la loi *a posteriori* pour les paramètres  $(p_l, q_l, m_l)$ . Les dépendances entre paramètres et données pour les processus de croissance et mortalité sont représentées sur la figure (3.5).

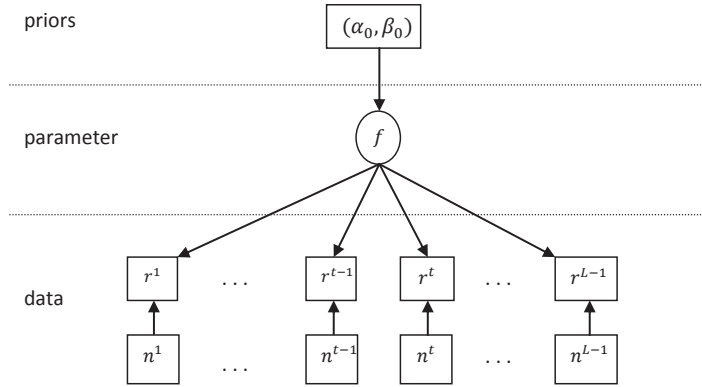


FIGURE 3.4 – DAG de processus de recrutement

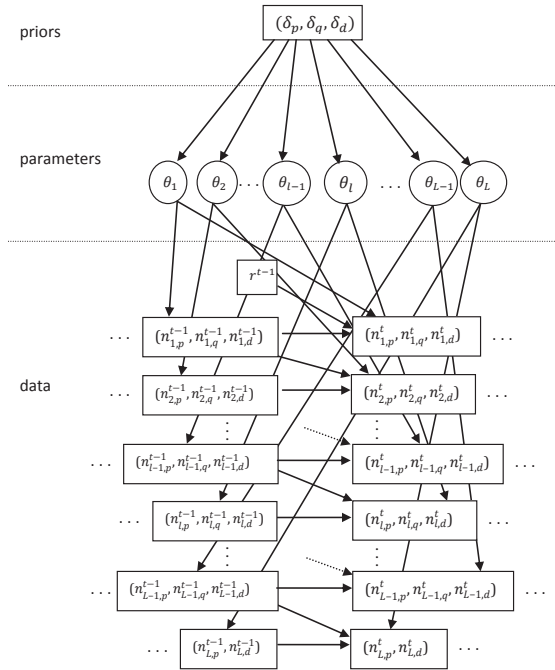


FIGURE 3.5 – DAG des processus de mortalité et croissance

## Mélanges de modèles d’Usher stochastique

Nous nous plaçons à présent dans le contexte d’une communauté de  $S$  espèces. Pour chaque espèce  $s$  de la communauté avec  $s = 1, \dots, S$ , nous avons les données  $Data_{t|t-1}^s$  permettant d’inférer le modèle d’Usher stochastique. Selon la théorie des groupes fonctionnels, il existerait un nombre inconnu de groupes  $K < S$  de dynamiques différentes au sein de la communauté. C’est à dire qu’il existe  $K$  modèles d’Usher stochastiques différents définis par  $K$  jeux de paramètres  $\theta^1, \dots, \theta^K$ . Ainsi, il faut déterminer à quel groupe appartient chaque espèce. Notons  $\vec{C} = (C_1, \dots, C_S)$  avec  $C_i \in \{1, \dots, K\}$  le vecteur des affectations des espèces dans les groupes. Par exemple  $C_2 = 3$  signifie que l’espèce 2 est affectée dans le groupe 3.

Pour construire les groupes fonctionnels en se basant sur les modèles d’Usher stochastiques, il suffit donc d’estimer  $K, \vec{C}$  et  $\theta^1, \dots, \theta^K$ . Par la formule de Bayes, nous avons une expression proportionnelle à la loi *a posteriori* des paramètres :

$$\pi(K, \vec{C}, \theta^{1:K} | Data_{t|t-1}^{1:S}) \propto \prod_{s=1}^S \mathcal{L}(Data_{t|t-1}^s | \theta^{C_s}) \pi_{\vec{C}|\vec{\theta}, K}^0(\vec{C} | \vec{\theta}, K) \pi_{\vec{\theta}|K}^0(\vec{\theta} | K) \pi_K^0(K) \quad (3.4)$$

avec  $\pi_{\vec{C}|\vec{\theta}, K}^0$ ,  $\pi_{\vec{\theta}|K}^0$  et  $\pi_K^0$  les loi *a priori* sur les paramètres inconnus.

Maximiser la loi *a posteriori* de l’équation (3.4) est une question difficile car le nombre de groupes  $K$  étant inconnu, les nombre de paramètres à estimer dans le modèle change avec les valeurs de  $K$ . Il n’est donc pas possible d’utiliser un algorithme MCMC pour estimer cette loi *a posteriori*. Une approche consiste à inférer le modèle conditionnellement à  $K$  en essayant toutes les valeurs de  $K = 1, \dots, S$  et de retenir la meilleure configuration selon un critère de sélection de modèle tel que AIC, BIC, DIC . . . nous l’avons explorée dans [Picard et al. \(2010\)](#). En statistique fréquentiste cette approche peut s’appuyer sur l’algorithme EM (Expectation-Maximisation) qui permet d’optimiser simultanément une classification et ses paramètres associés ([McLachlan and Krishnan, 2007](#))

Pour éviter que la classification repose sur le choix d’un critère de sélection de modèle, nous avons développé un algorithme MCMC qui autorise le nombre de paramètres inconnus à changer au cours de l’inférence. Il permet ainsi d’estimer  $K$  en même temps que les autres paramètres. Cet algorithme est du type Reversible Jump MCMC ([Richardson and Green, 1997](#)), sa description est détaillée dans notre article ([Mortier et al., 2013](#)). Il consiste à proposer des sauts sur le nombre de groupes : création d’un nouveau groupe (par scission d’un groupe en deux),



suppression d'un groupe (par fusion de deux groupes) ou conservation du même nombre de groupes en réaffectant les espèces. L'étape d'acceptation-rejet de la modification du nombre de groupe est faite avec ratio de Metropolis -Hastings modifié par rapport aux MCMC usuels car il intègre la probabilité de saut aillant mené à la modification du nombre de groupes.

Nous avons appliqué cette méthode de classification à des données issues du dispositif de Paracou ([http://antilles-guyane.cirad.fr/aux\\_antilles\\_et\\_en\\_guyane/guyane/site\\_de\\_paracou](http://antilles-guyane.cirad.fr/aux_antilles_et_en_guyane/guyane/site_de_paracou)), en Guyane française. Les groupes obtenus étaient cohérents avec les classifications d'espèces construites sur ce site (Favrichon, 1994; Gourlet-Fleury, 1997; Gourlet-Fleury and Houllier, 2000) et avec les comportements écologiques connus : les espèces de petites statures, hors pionnières, ont tendance à avoir une croissance lente et un turn-over élevé alors que les espèces de grande stature ont tendance à avoir une croissance rapide avec un turn-over lent. Ce résultat est logique car il a été montré dans Picard et al. (2012) que les mélanges de modèles d'Usher classaient les espèces selon leur stature et leur taux de croissance.

Nous avons formalisé dans le cadre bayésien et développé un algorithme regroupant automatiquement les espèces suivant la similarité de leurs dynamiques sans intervention de l'utilisateur. Cet algorithme a produit des classifications cohérentes, il constitue ainsi une contribution intéressante pour la communauté des écologues. Cependant, il est complexe à implémenter et peut être très long à converger. Donc même avec son code à disposition, il reste difficilement accessible à des personnes n'ayant une bonne maîtrise de la statistique bayésienne. Par ailleurs, son avantage conceptuel de ne pas reposer sur un critère de sélection de modèle n'est pas totalement exact si on y regarde de plus près. La plupart des critères de sélection de modèle consistent à pénaliser la vraisemblance suivant la complexité du modèle or le Reversible Jump MCMC fait un peu la même chose. En effet, la modification appliquée au ratio de Métropolis-Hastings pénalise la vraisemblance suivant la complexité induite par la modification du nombre de groupes. Ainsi, si on change la façon de créer ou supprimer des groupes on risque de ne pas obtenir la même classification.

Etant donnés tous ces éléments, il n'est pas possible de conclure à la supériorité de cette méthode de classification sur les autres.

### 3.2.2 Approche par les traits fonctionnels

L'approche par groupes fonctionnels présente aussi une faiblesse conceptuelle, la notion de groupes d'espèces est une vue de l'esprit. Beaucoup d'écologues croient en des continuums d'espèces sur un ou plusieurs gradients comportementaux. En particulier, il a été montré que les traits fonctionnels des espèces caractérisent ces gradients (Diaz et al., 2004). Mes collègues écologues d'Ecofog adhéraient plutôt à ce concept de continuums, nous avons donc cherché à relier les traits fonctionnels des espèces aux paramètres des processus de la dynamique forestière. Pour ce faire, nous avons réalisé des modèles bayésiens hiérarchiques, pour la croissance (Hérault et al., 2011) et pour la mortalité (Aubry-Kientz et al., [under review](#)) des arbres. La construction de ces deux modèles est détaillée dans les paragraphes suivants.

#### Modèle de croissance à l'échelle de la communauté intégrant les traits fonctionnels

La modélisation de la croissance des arbres en diamètre a été largement étudiée dans la littérature, on peut consulter une synthèse des modèles les plus courants dans Franc et al. (2000). Ce sont des modèles ontogéniques, c'est à dire que l'accroissement est a minima fonction du diamètre courant de l'arbre. Il est intéressant de remarquer que deux points de vue existent sur les propriétés générales d'un modèle de croissance en diamètre : soit l'arbre s'arrête de croître à partir d'un certain diamètre, soit l'arbre continue de croître jusqu'à sa mort mais de moins en moins. La deuxième possibilité me semble plus réaliste si l'on s'en tient à la structure des troncs d'arbres. En effet, l'anneau cambial constituant l'extrémité du tronc ne peut que croître tant que l'arbre est vivant. Ceci dit, il est tout à fait envisageable que les mesures de diamètre, au mètre ruban, sur les gros arbres ne permettent pas d'observer de faibles accroissements alors qu'ils sont réels au niveau cellulaire.

Dans notre cas, nous avons étudié la croissance des arbres des 50 espèces présentes sur les 6 parcelles témoins de Paracou ayant plus de 13 individus vivants entre 1991 et 2007. Nous avons cherché à prédire la croissance observée sur cette période de 16 ans, comme il s'agit d'arbres en parcelles témoins nous avons négligé l'effet de la compétition dans le modèle. Avant d'essayer de remplacer les espèces par leurs traits, nous avons comparé différents modèles sur chacune des espèces afin d'en trouver un qui convienne au plus grand nombre : Gompertz, Korf, Logistique, Weibull (cf (Zeide, 1993), celui utilisé dans (Canham et al., 2004; Uriarte et al., 2004) et une fonction puissance.

Nous avons retenu le modèle utilisé dans (Canham et al., 2004; Uriarte et al., 2004) défini par l'équation (3.5) car il avait le meilleur AIC :

$$\log(Ac + 1) = Gmax \times \exp\left\{-\frac{1}{2}\left(\frac{\log\left(\frac{dbh}{Dopt}\right)}{K}\right)\right\} + \varepsilon \quad (3.5)$$

avec  $Ac$  l'accroissement de l'arbre sur la période observée,  $dbh$  le diamètre de l'arbre au début de la période d'observation,  $Gmax$ ,  $Dopt$ ,  $K$  les paramètres du modèle et  $\varepsilon$  l'erreur résiduelle. Ce modèle ne suppose pas que la croissance des arbres s'arrête à partir d'un certain diamètre et ses paramètres ont une interprétation biologique simple (cf. figure 3.6). En effet,  $Gmax$  correspond à l'accroissement maximal que l'arbre atteint lorsque son diamètre est égal à  $Dopt$ .  $K$  indique si la courbe de croissance en fonction du diamètre est plus ou moins piquée autour de ce maximum. Pour chacune des 50 espèces, nous avons calibré le triplet de paramètres  $(Gmax_s, Dopt_s, K_s)$  avec  $s = 1, \dots, 50$ . Il est assez intuitif de relier le tempérament des espèces aux valeurs des trois paramètres, comme nous avons pu le vérifier sur les 50 espèces de notre étude. Par exemple si on s'intéresse aux tempéraments extrêmes, les espèces héliophiles pionnières vont avoir relativement aux autres espèces un  $Gmax$  grand, un  $Dopt$  petit et un  $K$  petit illustrant leur capacité à croître rapidement dès qu'une ouverture se présente sans atteindre la canopée. Les espèces sciaphiles de sous-bois vont avoir un  $Gmax$  petit, un  $Dopt$  petit et un  $K$  grand illustrant une croissance faible tout au long de la vie de l'arbre et une taille petite. Les espèces de canopée vont avoir un  $Gmax$  plutôt grand, un  $Dopt$  grand et un  $K$  moyen.

Le fait que les paramètres s'interprètent biologiquement est d'autant plus intéressant par rapport à notre objectif d'introduire des traits fonctionnels à la place de l'espèce. Ainsi nous avons remplacé chacun des paramètres par une combinaison linéaire des traits fonctionnels. Nous disposons de 17 traits fonctionnels mesurés dans le cadre du projet Bridge (cf. table 3.1).

Si on note  $T_i^j$  la valeur du trait  $j$  pour l'arbre  $i$ , le modèle unique pour n'importe quel arbre  $i$ , quelque soit son espèce, serait :

$$\log(Ac_i + 1) = (\theta_G^0 + \sum_{j=1}^{17} \theta_G^j T_i^j) \times \exp\left\{-\frac{1}{2}\left(\frac{\log\left(\frac{dbh_i}{\theta_D^0 + \sum_{j=1}^{17} \theta_D^j T_i^j}\right)}{\theta_K^0 + \sum_{j=1}^{17} \theta_K^j T_i^j}\right)\right\} + \varepsilon_i \quad (3.6)$$

Etant donné les corrélations existantes entre les traits, afin de pouvoir interpréter biologiquement le rôle des traits dans le modèle nous les avons sélectionnés. Comme il est numériquement très lourd d'explorer toutes les combinaisons de

	Units	Role
Leaf economics		
Foliar $\delta^{13}\text{C}$ composition ( $\delta^{13}\text{C}$ ) <sup>a</sup>	‰	Resource capture
Foliar C:N (C/N) <sup>a</sup>	cg g <sup>-1</sup>	Resource capture and defence
Foliar Km (K) <sup>a</sup>	mg g <sup>-1</sup>	Resource capture
Foliar Nm (N) <sup>a</sup>	cg g <sup>-1</sup>	Resource capture
Foliar Pm (P) <sup>a</sup>	mg g <sup>-1</sup>	Resource capture
Leaf tissue density (LTD) <sup>a</sup>	g cm <sup>-3</sup>	Resource capture and defence
Laminar total chlorophyll (chlorophyll) <sup>a</sup>	µg mm <sup>-2</sup>	Resource capture
Laminar toughness (toughness) <sup>a</sup>	N	Defence
Specific leaf area (SLA) <sup>a</sup>	cm <sup>2</sup> g <sup>-1</sup>	Resource capture and defence
Life history		
Forest layer (Layer) <sup>b</sup>	Understorey, canopy or emergent	Architecture
Maximum height (Height) <sup>a</sup>	m	Architecture and structure
Maximum diameter (d.b.h.max) <sup>a</sup>	mm	Architecture and structure
Seed mass <sup>a</sup>	g	Reproductive investment
Stem and branch orientation <sup>b</sup>	Orthotropic or plagiotropic	Architecture
Stem economics		
Trunk bark thickness (Bark) <sup>a</sup>	mm	Defence
Trunk wood moisture content (Wood moisture) <sup>a</sup>	%	Transport and structure
Trunk xylem density (Wood density) <sup>a</sup>	g cm <sup>-3</sup>	Transport, structure and defence

TABLE 3.1 – Traits fonctionnels utilisés pour prédire la croissance des arbres (extrait de [Hérault et al. \(2011\)](#))

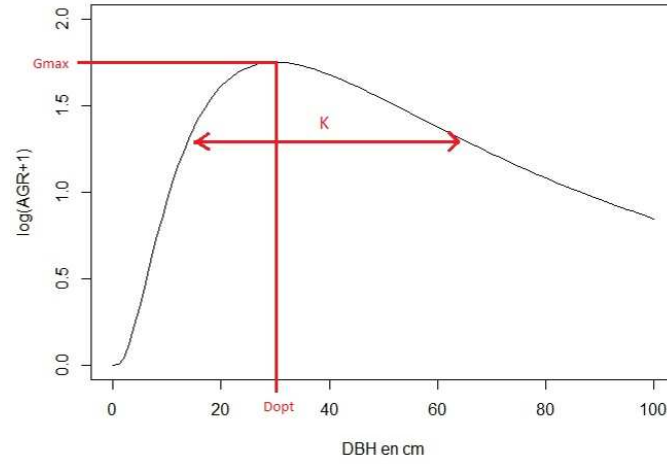


FIGURE 3.6 – Graphique du modèle de croissance (3.5)

traits pour les trois paramètres dans le modèle (3.6), nous avons sélectionné les variables avant d'inférer le modèle. Pour chacun des paramètres  $Gmax$ ,  $Dopt$  et  $K$  nous avons réalisé une régression contre les valeurs des traits des espèces et conservé l'ensemble des traits qui maximisait le BIC. C'est à dire, si on prend le cas de  $Gmax$ , pour chacune des espèces nous avons une valeur  $Gmax_s$ , ainsi que les valeurs des traits  $T_s^1, \dots, T_s^{17}$  avec  $s = 1, \dots, 50$ . Nous avons réalisé la régression linéaire  $Gmax_s = \sum_j \beta_j T_s^j + \varepsilon_s$  et retenu la combinaison de traits optimisant le BIC. Ce qui nous a conduit au modèle de croissance :

$$\begin{aligned} \log(Ac_i + 1) = & (\theta_G^1 wsg + \theta_G^2 Dbh.max + \theta_G^3 Height + \theta_G^4 \delta^{13}C) \\ & \times \exp\left\{-\frac{1}{2}\left(\frac{\log(\frac{dbh_i}{\theta_D Dbh.max})}{\theta_K wsg}\right)^2\right\} + \varepsilon_i \end{aligned} \quad (3.7)$$

avec  $wsg$  la densité du bois de l'espèce,  $Dbh.max$  le 95<sup>ème</sup> percentile du diamètre des arbres de l'espèce,  $Height$  la hauteur maximale des arbres de l'espèce,  $\delta^{13}C$  la signature isotopique des feuilles de l'espèce et  $\varepsilon_i$  l'erreur supposée suivre une loi normale centrée.

Nous avons réalisé l'inférence du modèle (3.7) avec un algorithme de Metropolis-within Gibbs. Car conditionnellement aux paramètres  $\theta_D$  et  $\theta_K$  le modèle est quasi-linéaire il est ainsi possible d'obtenir la loi marginale *a posteriori* pour les paramètres  $\theta_G^1, \theta_G^2, \theta_G^3, \theta_G^4$ . Nous avons ainsi obtenu le premier modèle de crois-

sance à l'échelle de la communauté en forêt tropicale basé sur les traits fonctionnels.

Pour évaluer la performance du modèle, nous avons comparé trois niveaux d'agrégation des arbres : (i) tout le peuplement, soit les mêmes  $G_{max}$ ,  $D_{opt}$  et  $K$  pour tous les arbres avec au total 4 paramètres à estimer ; (ii) le modèle avec les traits avec 7 paramètres à estimer (iii) un modèle par espèce avec 200 paramètres à estimer. En terme de BICM (adaptation du BIC pour les modèles bayésiens ([Raftery et al., 2007](#))), notre modèle avec les traits est le meilleur compromis. Pour avoir une idée plus précise de la perte en passant des espèces aux traits, les RMSEP des trois modèles sont 57, 51 et 43 mm/an. Grossièrement on peut dire que les traits permettent de capter quasiment la moitié de la variabilité expliquée par les espèces. Le gain en nombre de paramètres étant très important ce compromis est donc vraiment intéressant.

### **Modèle de mortalité à l'échelle de la communauté intégrant les traits fonctionnels**

Les résultats obtenus avec l'utilisation des traits pour faire un modèle de croissance à l'échelle de la communauté étant convaincants, nous avons décidé de faire un sujet de thèse pour généraliser cette approche aux autres processus de la dynamique forestière, la mortalité et le recrutement. L'objectif étant d'intégrer ces modèles dans un simulateur de forêt tropicale arbre centré pour être en mesure de tester la réponse de la communauté aux différents scénarios climatiques. Cette thèse réalisée par Mélaïne Aubry-Kientz a commencée en octobre 2011, elle a logiquement débutée par la construction d'un modèle de mortalité à l'échelle de la communauté basé sur les traits fonctionnels. Ce travail est présenté dans l'article [Aubry-Kientz et al. \(under review\)](#).

Le processus de mortalité a été beaucoup moins étudié que la croissance pour plusieurs raisons. D'une part c'est un phénomène complexe car il a plusieurs causes : sénescence, maladie, manque de ressource (eau ou lumière), chablis (chute) ou accident (causé par la chute d'un autre arbre). D'autre part, c'est un phénomène ponctuel relativement rare qui concerne à peu près de 2% de la population sur un an en forêt tropicale humide ([Condit et al., 1995](#)) et environ 1% en dans les forêts guyanaises. Si on rajoute la grande diversité spécifique, il devient difficile de modéliser ce processus pour chaque espèce. Cela nécessiterait des suivis de grandes parcelles sur une très longue période pour récupérer suffisamment d'information. Différentes approches ont été proposées : taux de mortalité par classe de diamètre

(Bohlman and Pacala, 2012) ou par groupes d'espèces (Alder and Silva, 2000), modèle de mortalité individuel faisant intervenir les facteurs environnementaux (densité du voisinage, type de sol, . . .) et des caractéristiques individuelles (DBH, croissance passée, . . .) (Phillips et al., 2004b; Gourlet-Fleury et al., 2005). La grande diversité d'espèces pose toujours le problème d'un nombre trop important de paramètres à estimer par rapport au nombre de données à disposition. Nous avons opté par la même stratégie que dans (Hérault et al., 2011) en utilisant des traits fonctionnels pour expliquer les différences spécifiques de la mortalité. Mais nous avons utilisé seulement 15 des 17 traits du tableau 3.1 car 2 traits sur le bois qui étaient renseignés pour peu d'espèces, ont été omis de cette étude. Nous avons également utilisé les données des 6 parcelles témoins de Paracou de 1992 à 2010. Sur les 20408 arbres recensés en 1992, 17450 étaient encore vivants en 2010. Nous avons dû gérer deux difficultés dues à la nature de ces données. Tout d'abord la détermination botanique n'ayant été complétée sur les parcelles témoins qu'en 2011, 85% des arbres morts sur la période n'ont pas été identifiés et possèdent seulement un nom vernaculaire. Il s'agit d'un nom donné par la population locale qui discrimine généralement les arbres selon l'usage de leur bois. Ensuite, les valeurs des 15 traits sont connus pour seulement 51% des arbres et 20% des 335 espèces présentes.

Pour gérer rigoureusement les incertitudes induites par ces informations manquantes, nous avons construit des modèles de connaissance. Pour résumer, il y a trois cas de figures où les données sont incomplètes ou manquantes :

1. l'arbre est identifié à l'espèce mais un ou plusieurs traits ne sont pas renseignés
2. l'arbre n'est pas déterminé à l'espèce mais seulement au genre ou à la famille
3. l'arbre est mort avant d'avoir été identifié, seulement son nom vernaculaire est disponible

Nous avons géré les deux premiers cas avec le même modèle. Comme les traits fonctionnels utilisés sont bien conservés phylogénétiquement (Baraloto et al., 2012) nous utilisons l'information d'appartenance au même genre/famille pour modéliser la valeur des traits. Plus précisément pour un trait  $T$ , soit  $\{s_1, \dots, s_{n_T}\}$  l'ensemble des  $n_T$  espèces de la communauté pour lesquelles nous avons des valeurs associées  $\{t_1, \dots, t_{n_T}\}$  pour ce trait  $T$ . Pour tout arbre d'une espèce  $s \in \{s_1, \dots, s_{n_T}\}$ , la valeur de son trait est  $t_s$ . Pour tout arbre  $a$  d'une autre espèce, la valeur du trait est modélisé par une loi multinomiale  $\mathcal{M}((\tilde{t}_1, \dots, \tilde{t}_{g_T}), (\nu_1, \dots, \nu_{g_T}))$ .

Où  $(\tilde{t}_1, \dots, \tilde{t}_{g_T})$  est un sous-ensemble de  $\{t_1, \dots, t_{n_T}\}$  pour lesquelles les espèces associées sont du même genre/famille que celui de l'arbre  $a$  et  $\nu_i = \frac{N_i}{N}$ ,  $N_i$  le nombre d'arbre du même genre/famille que l'arbre  $a$  avec leur trait égal à  $\tilde{t}_i$ ;  $N$  le nombre total d'arbre du même genre/famille que l'arbre  $a$ .

Nous avons géré le troisième cas avec un modèle bayésien. Le principe général est le même que précédemment, pour tout arbre ayant un nom vernaculaire  $v$ , la valeur de son trait est supposée suivre une distribution multinomiale  $\mathcal{M}((t_1, \dots, t_{n_T}), (\alpha_1, \dots, \alpha_{n_T}))$ . Les probabilités  $\alpha^v = (\alpha_1, \dots, \alpha_{n_T})$  ont été déterminées avec un modèle bayésien de la relation entre les noms vernaculaire et les noms latin des espèces. Nous avons utilisé deux types d'information : l'expertise des prospecteurs et les données des inventaires récents :

**L'expertise des prospecteur** a été utilisée comme information *a priori* elle renseigne sur les espèces pouvant être associées à un nom vernaculaire donné. Pour un nom vernaculaire  $v$ , cela se formalise par un vecteur  $\lambda^v = (\lambda_1, \dots, \lambda_{n_T})$  où  $\lambda_i = \frac{1}{m_v}$  si les arbres de l'espèce  $s_i$  peuvent se faire attribuer le nom vernaculaire  $v$  avec  $m_v$  le nombre d'espèces associées par les experts au nom vernaculaire  $v$  et  $\lambda_i = \frac{\varepsilon}{n-m_v}$  si les arbres de l'espèce  $s_i$  ne peuvent pas se faire attribuer le nom vernaculaire  $v$  avec  $\varepsilon$  fixé petit arbitrairement pour intégrer les erreurs potentielles.

**Les données d'inventaire** ont été utilisées pour mettre à jour cette information elles renseignent sur les fréquences d'espèces observées pour le nom vernaculaire. Pour un nom vernaculaire  $v$ , cela se formalise par un vecteur  $f^v = (f_1, \dots, f_{n_T})$  où  $f_i$  est le nombre d'arbre avec le nom vernaculaire  $v$  ayant été déterminés de l'espèce  $i$ .

**Le couplage** de l'expertise  $\lambda^v$  et des données  $f^v$  a été réalisé avec un modèle Multinomial-Dirichlet (Bernardo and Smith, 1994). L'information d'expertise a été utilisé comme hyper-paramètres de la loi *a priori* Dirichlet de  $\alpha^v$ ,  $\pi_{\alpha^v}^0 = \mathcal{D}(\lambda^v)$ . Nous avons modélisé les valeurs de  $f^v$  par une distribution multinomiale de paramètres  $\alpha^v$ . Ainsi comme les lois Dirichlet et multinomiale sont conjuguées (Bernardo and Smith, 1994) la loi *a posteriori* de  $\alpha^v$  est aussi une loi de Dirichlet,  $\pi_{\alpha^v|f^v} = \mathcal{D}(\lambda^v + f^v)$ .

Comme l'incertitude sur les traits fonctionnels est formalisée par les modèles précédents, il est possible d'intégrer les traits comme des variables aléatoires dans un modèle de mortalité. Nous avons utilisé un modèle linéaire généralisé avec un lien *logit* (McCullagh and Nelder, 1989) pour expliquer qu'un arbre soit mort ou



soit resté vivant entre 1992 et 2010. En plus des 15 traits, nous avons ajouté deux variables ontogéniques dans le modèles  $\frac{dbh}{DBH_{max}}$  et  $(\frac{dbh}{DBH_{max}})^2$  avec  $DBH_{max}$  le 95<sup>ème</sup> percentile des diamètres observés des espèces.

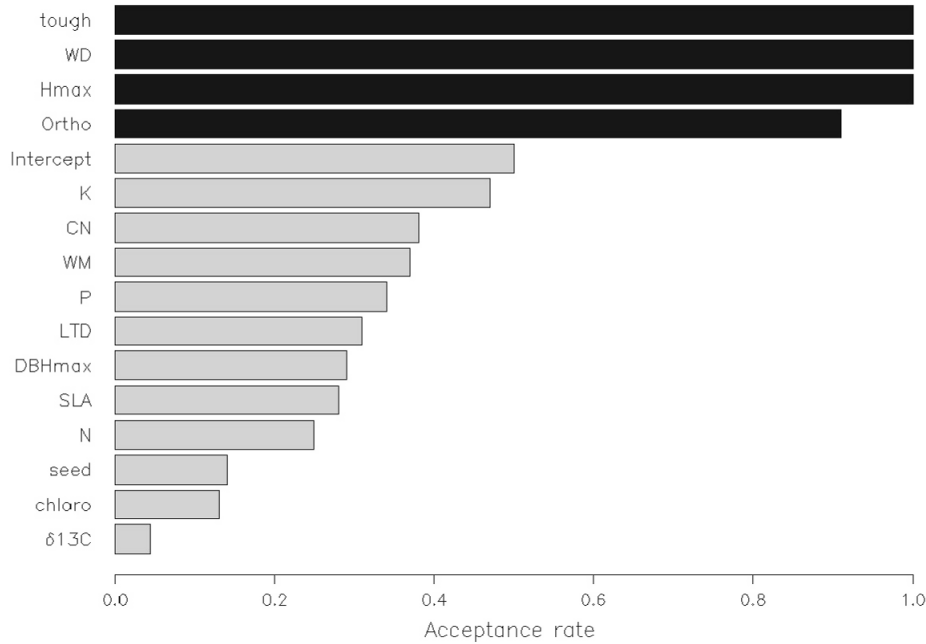


FIGURE 3.7 – Sélection de variables aléatoire par Kuo-Mallick (extrait de (Aubry-Kientz et al., under review))

Comme pour le modèle de croissance, il faut réaliser une sélection de variables pour être en mesure de les interpréter biologiquement. Or comme les traits sont aléatoires cela complique les choses. Nous avons opté pour adapter l'algorithme de sélection de Kuo-Mallick (Kuo and Mallick, 1998). Le principe de cet algorithme de type MCMC est de multiplier chacune des variables, en plus de leurs paramètres respectifs, par une indicatrice (variable aléatoire valant 0 ou 1). Au fil des itérations de l'algorithme les indicatrices sont mises à jour comme tous les autres paramètres. Lorsque l'algorithme atteint son état stationnaire, on peut choisir les variables à garder dans le modèle à partir des fréquences des indicatrices. Dans notre cas, comme les variables sont aléatoires, à chaque itération de l'algorithme nous avons tiré de nouvelles valeurs pour les traits des arbres suivant les modèles Multinomiaux-Dirichlet décrits précédemment. Ainsi l'incertitude des variables explicatives était prise en compte dans le processus de sélection. Comme

on peut le voir sur la figure 3.7 cette procédure de sélection a bien discriminé les variables, nous avons choisi les variables à garder selon l’heuristique du coude. Nous avons ainsi construit un modèle de mortalité à l’échelle de la communauté basé sur les traits fonctionnels, aillant un nombre de paramètres faible relativement à la complexité de processus modélisé. Par ailleurs, la méthode que nous avons développée pour prendre en compte rigoureusement les incertitudes de détermination est originale. Elle sera certainement transposable dans d’autre étude car il est très fréquent que les déterminations soit incomplètes en écologie tropicale.

### 3.2.3 Groupes fonctionnels versus Traits fonctionnels

Les groupes fonctionnels que nous avons construits relèvent du cadre fondamental de [Lavorel et al. \(1997\)](#). Cette théorie vise à classer les espèces dans un nombre de groupes restreints. La diversité de la communauté serait ainsi plus facile à appréhender puisqu’il suffirait d’analyser le fonctionnement de chaque groupe et les interactions entre groupes. Autrement dit les espèces d’un groupe sont équivalentes au sens de la caractéristique d’intérêt, ainsi les nuances entre espèces, en dessous d’un certain seuil, sont négligées. La diversité est donc modélisée par des barycentres locaux des espèces, minimisant un compromis entre la variance intra et la variance inter groupes de la caractéristique d’intérêt.

L’approche par les traits fonctionnels est conceptuellement différente puisque l’hypothèse sous-jacente est qu’il existe un continuum dans les stratégies de fonctionnement des espèces. Il n’est pas question de réduire la diversité en classant les espèces dans des groupes mais d’appréhender ce continuum. Les traits fonctionnels sont utilisés pour représenter les espèces dans ce continuum. La diversité est donc modélisée par un nuage de points, associés aux espèces, dans l’espace engendré par les traits fonctionnels.

Les deux conceptions ne sont pas à mettre en opposition car parmi les méthodes de classification en groupes : fonctionnels, émergents, de stratégies ou de réponses ([Lavorel et al., 1997](#)), la plupart s’appuyaient sur des traits. La notion de trait n’était pas encore stabilisée et la modélisation hiérarchique permettant d’injecter les traits dans les modèles de dynamiques n’était pas encore répandue en écologie. Aussi je vois plutôt l’approche par les traits fonctionnels comme le prolongement naturel de l’approche par groupes fonctionnels suite à l’émergence des grandes bases de données de traits ([Chave et al., 2009](#); [Baraloto et al., 2010](#)) et à la dissémination de la statistique bayésienne en écologie.

Concernant les avantages et inconvénients de ces deux approches, je présente mon analyse sous l'angle théorique puis sous l'angle pratique.

Théoriquement, l'approche par traits fonctionnels me semble plus satisfaisante car elle tend à préserver la diversité des fonctionnements des espèces. Cependant, elle nécessite d'une part que les traits mesurés soient les plus pertinents pour expliquer les différences de fonctionnement des espèces ; et d'autre part, que le lien entre les traits et le fonctionnement soit identifié ou à défaut linéaire. Pour préciser ce deuxième point, considérons le cas du modèle de mortalité de la section 3.2.2 : nous avons introduit une variable relative au stade ontogénique de l'espèce,  $dbh/dbh.max$ . Or il est reconnu que le taux de mortalité n'augmente pas linéairement avec le stade ontogénique mais que le lien a plutôt une forme en U. Ce qui exprime une mortalité élevée au jeune stade qui diminue rapidement pour se stabiliser à un niveau faible et qui ré-augmente lentement lorsque l'individu passe au stade de sénescence. Nous avons ainsi ajouté le terme de deuxième degré  $(dbh/dbh.max)^2$  pour être en mesure de se rapprocher de cette forme en U.

Mais généralement, la forme des liens entre les traits et les processus n'est pas connue. Par défaut, nous avons pris des liens linéaires. Il est tout à fait possible qu'en procédant ainsi, des traits ne soient pas retenus ou perdent de leurs pouvoirs explicatifs car le lien était de type uni-modal ou plus largement non linéaire. De plus, il faut choisir un critère de sélection de modèle pour identifier les traits à garder dans le modèle.

L'approche par groupes fonctionnels n'est pas soumise à des choix sur des formes de modèle, mais uniquement au choix d'un critère de sélection de modèles pour définir le nombre de groupes. Cependant, le nombre de groupes retenu ne va pas varier beaucoup entre les critères de sélection. Ainsi, cette approche est plus robuste, dans le sens où les choix de l'utilisateur ne peuvent pas beaucoup influencer les résultats. De plus, sous l'hypothèse du continuum d'espèces, l'approche par groupes consisterait à partitionner ce continuum ce qui peut être tout à fait pertinent si le continuum n'est pas trop dispersé.

En pratique, de par mes travaux, je n'ai pas acquis un avis tranché sur l'approche à privilégier, groupes fonctionnels ou traits fonctionnels. Principalement, car je n'ai pas encore utilisé ces deux approches sur un même modèle et un même jeu de données : j'ai utilisé les groupes fonctionnels sur des modèles matriciels et j'ai utilisé les traits fonctionnels sur modèles de croissance ou mortalité à l'échelle de l'arbre. Les deux approches m'ont permis d'atteindre l'objectif de modéliser la dynamique à l'échelle de la communauté, mon analyse va donc se focaliser sur la

mise en œuvre de ces approches.

Pour les cas où les traits des arbres sont complètement connus, l'approche avec les traits se met en œuvre aisément, puisqu'il «suffit» de calibrer le modèle pour chaque combinaison de traits et retenir celle qui maximise le critère de sélection utilisé. Si le modèle est complexe, il en sera de même de son optimisation et s'il y a beaucoup de traits l'exploration des combinaisons demandera du temps de calcul. Mais si l'incertitude sur les traits est intégrée, la situation se complique considérablement, il faut alors avoir recours à une approche comme celle que nous avons développée pour le modèle de mortalité. En d'autres termes, il faut développer un algorithme d'inférence spécifique et le temps de calcul nécessaire peut être très élevé.

L'approche par groupes fonctionnels se situerait à mon avis entre les deux en terme de complexité. Cette fois il s'agit d'explorer des combinaisons d'espèces à un nombre de groupes donné, en faisant varier ce nombre de groupes 1 au nombre d'espèces de l'échantillon. C'est le nombre de groupes et sa combinaison d'espèces associée maximisant le critère de sélection qui est retenu. Il existe des *packages* sous R, par exemple utilisant l'algorithme Expectation-Maximization (EM), permettant de faire quasi-automatiquement cette procédure.

En ce qui concerne le matériel nécessaire, les différences sont plus marquées. L'approche par groupes fonctionnels nécessite seulement la détermination botanique des arbres et il est même possible de l'utiliser avec une détermination simplement au niveau vernaculaire. L'approche par les traits nécessite de connaître les traits de tous les arbres ce qui généralement se réduit en pratique à connaître les traits de toutes les espèces des arbres présents. Autrement dit, il faut avoir une détermination botanique complète. Donc pour la plupart des jeux de données actuels, en toute rigueur, il faudrait intégrer l'incertitude sur les traits ce qui nécessiterait d'avoir recours à la méthode d'inférence la plus complexe.

Pour résumer mon opinion, l'approche par groupes fonctionnels est opérationnelle pour certaines questions et est bien adaptée aux données d'inventaires traditionnels. L'approche par traits fonctionnels va à terme la supplanter car elle appréhende d'une façon bien adaptée la grande diversité spécifique des forêts tropicales. Cela permet de s'affranchir de la notion d'espèces qui finalement n'apporte pas l'information adéquate pour construire des modèles dans un milieu si diversifié. Cependant, la question de l'homogénéité des protocoles de mesures des traits entre études se pose, dans un contexte où la réalisation de méta-analyses est de plus en plus fréquente. Une fois que ces points seront stabilisés, il me paraît vraisemblable que des outils de mesures de traits pour le terrain seront élaborés

et qu'ils seront mesurés en routine lors des inventaires. Il est même imaginable qu'à terme ces données de traits remplaceront les déterminations botaniques puisqu'elles apportent beaucoup plus d'information sur le peuplement et sur son fonctionnement.

### **3.3 Vers une quantification de l'impact des changements climatiques sur les forêts tropicales**

Selon les prévisions des modèles climatiques (IPCC, 2007) les forêts devraient être soumises à des situations climatiques inédites. Explorer les impacts sur les forêt et les effets rétro-actifs est un enjeu pour les scientifiques qui ne pourra être abordé que par des recherches pluridisciplinaires (Bonan, 2008). Nous avons contribué à cette grande question en étudiant les capacités de résilience des forêts tropicales humides face à ces changements climatiques attendus. Nos études se sont focalisées sur la Guyane. Dans les zones tropicales, il est probable que les changements de régime pluviométrique auront un plus grand impact sur la végétation que l'augmentation de la température (Engelbrecht et al., 2007). En effet, les alternances régulières de saisons sèches et saisons des pluies typiques des forêts tropicales ont probablement façonné les communautés d'espèces. La méta-analyse de (Phillips et al., 2004a) a suggéré l'existence de changements dans les processus de dynamique forestière. Ces changements se caractérisent par une accélération du turn-over, c'est à dire augmentation des taux de recrutement et de mortalité. Comme des changements de régime pluviométrique ont déjà été observés dans différentes zones tropicales (Malhi et al., 2004) et que les modèles climatiques suggèrent une intensification de ces modifications pour le siècle prochain, il nous a paru important de développer des modèles pour prédire la réponse de la forêt tropicale humide aux différents scénarios climatiques envisagés. Nous avons construit une thèse pour avancer sur cette thématique, qui a été réalisée par Fabien Wagner.

Cependant, aucune année avec un déficit de pluviométrie élevé ou avec des perturbations fortes du régime des pluies n'a été observée en Guyane depuis que la croissance des arbres est suivie de façon régulière. Comme nous n'avions pas les moyens de mettre en place une expérimentation d'exclusion des précipitations (Nepstad et al., 2002) nous avons cherché à étudier les différences de réactions, en terme de dynamique, des espèces au sein de la communauté lors des stress

hydriques. De telles études ont été menées à grande échelle pour les forêts tempérées ([Pastor and Post, 1988](#)), beaucoup moins dans les tropiques ([Malhi et al., 2004](#)). Dans un premier temps, nous avons essayé de mettre en relation la croissance annuelle des espèces avec des variables climatiques dont la pluviométrie puisqu'elle devrait être prioritairement affectée par les changements climatiques. Ces analyses préliminaires n'ont pas été très probantes si ce n'est qu'elles nous ont permis d'identifier un point nécessitant d'être éclairci pour atteindre notre objectif.

Dans la littérature, il n'y a pas de consensus sur la façon d'intégrer la pluviométrie dans les modèles de croissance pour tester l'effet éventuel d'un stress hydrique. Et plus ennuyeux, il n'y avait pas de moyen en forêt tropicale pour déterminer si les arbres sont en situation de stress hydrique. En pratique les auteurs, utilisaient par exemple le nombre jour sans pluie dans l'année ou le nombre maximal de jours consécutifs sans pluie ou le nombre de jours où la pluviométrie est inférieur à un seuil dépendant de l'évapotranspiration . . . Le problème de ces indices, mis à part leur côté arbitraire, est qu'ils n'intègrent pas la dynamique du système. Par exemple, ne pas prendre en compte la capacité du sol à conserver l'eau et le volume des pluies entre deux périodes sèches entraîne que les indicateurs auront des valeurs proches pour des niveaux de stress hydrique différents pour les arbres.

Comme nous cherchions à analyser le lien entre les variations de croissance et les variations de pluviométrie, soit un signal *a priori* faible puisque les variations de pluviométries inter-annuelles sont faibles, nous devons lever ces ambiguïtés. Ainsi nous avons entrepris de développer un modèle pour estimer l'eau disponible dans le sol pour les arbres en fonction de la pluviométrie. Ce modèle du cycle de l'eau en forêt tropicale a constitué le cœur de la thèse de Fabien Wagner, il est présenté dans l'article [Wagner et al. \(2011\)](#). Notre modèle de bilan hydrique journalier s'inspire du modèle BILJOU de [Granier et al. \(1999\)](#) pour les forêts tempérées. Le principe général consiste à partitionner le système en trois compartiments atmosphère, forêt, sol et décrire les échanges d'eau entre les compartiments (cf. figure 3.8). Quant au compartiment sol, il est divisé en couches caractérisées par le volume de racine présente qui conditionne comment l'eau est extraite par les arbres. Le volume d'eau maximal disponible dans une couche du sol pour les arbres est défini par la différence entre la capacité au champ et le point de flétrissement permanent de la couche du sol. La capacité au champ est le volume d'eau maximal que peut contenir le sol, au delà l'eau ruisselle ou est drainée. Le point de flétrissement permanent est le volume minimal d'eau dans le sol en dessous

duquel les arbres ne peuvent plus extraire d'eau. Le modèle doit permettre d'estimer quotidiennement l'eau disponible et l'eau disponible relative ( $REW$  = eau disponible divisée par volume d'eau maximale disponible) pour les arbres dans chaque couche.

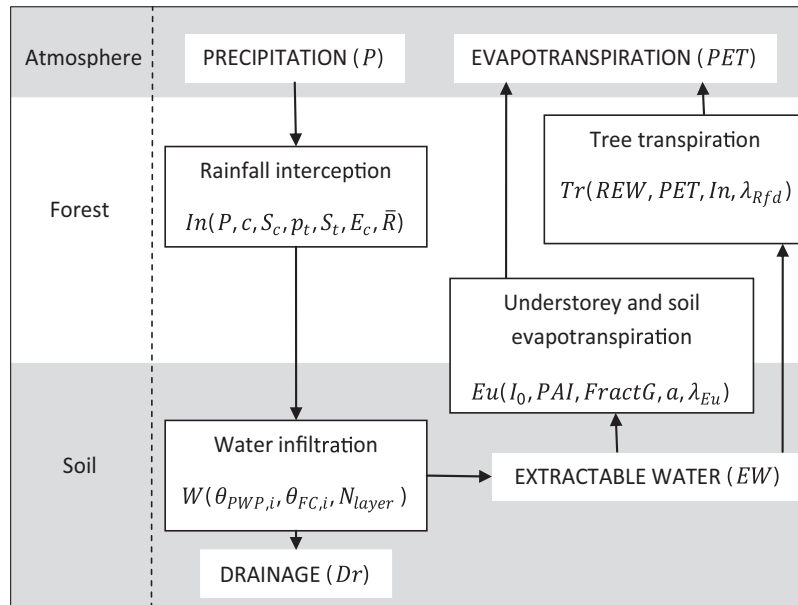


FIGURE 3.8 – Schéma du modèle de cycle de l'eau (extrait de [Wagner et al. \(2011\)](#))

Plus précisément les sous modèles présentés sur la figure (3.8) s'imbriquent de la façon suivante :

- La pluie provenant de l'atmosphère est en partie interceptée par les arbres, sous-modèle *Rainfall interception*. Cette quantité varie essentiellement suivant la structure de la canopée et de son niveau de saturation en eau
- La pluie non interceptée arrive au sol et s'infiltré, sous-modèle *Water infiltration*. La couche supérieure du sol se remplit jusqu'à sa capacité au champ, lorsqu'elle est pleine, s'il reste de l'eau, la couche suivante se remplit jusqu'à sa capacité au champ et ainsi de suite jusqu'à la dernière couche avec des racines. L'eau résiduelle sort du système par drainage vertical profond.
- L'eau disponible dans sol pour les arbres, *Extractable water*, est obtenue en faisant la somme de l'eau disponible dans chaque couche pondérée par la distribution des racines dans le sol.
- L'eau est extraite du sol par deux mécanismes d'évapotranspiration :

- sous-modèle *Understorey and soil evapotranspiration* : le volume d'eau évaporé est proportionnel à la quantité d'énergie arrivant au sol qui dépend de l'irradiance et de la structure du couvert de la canopée. L'eau est prise dans les couches supérieures du sol avec une intensité décroissant exponentiellement avec la profondeur.
- sous-modèle *Tree transpiration* : le volume d'eau transpiré par les arbres correspond à l'évapotranspiration potentielle quand l'eau dans le sol n'est pas limitante ( $REW > 0.4$ ), sinon pour  $REW < 0.4$  elle décroît linéairement suivant  $REW$ . Ce volume d'eau est extrait du sol selon la distribution des racines et les  $REW$  de chaque couche.

Nous avons calibré le modèle avec des données d'eau disponible dans le sol issues de sondes par électromécanique temporelle (TDR) et de pluviométrie mesurée dans le cadre du projet GuyaFlux (<http://www.ecofog.gf/spip.php?article365>) installé sur le site de Paracou. Ce modèle a constitué une première en milieu tropical. Notre travail a été au-delà d'un transfert en tropical d'un résultat obtenu en tempéré. En effet, nous avons fait plusieurs innovations méthodologiques :

- Le sous-modèle d'interception de la pluie a dû être remplacé par un modèle validé en forêt tropicale (Gash et al., 1995)
- Le découpage des couches de sol ne se fait plus par une approche empirique visant à être en adéquation avec la diminution de la proportion de racine suivant la profondeur. Nous découpons le sol en couches fines de 1 cm et la distribution des racines est modélisée par une loi exponentielle fonction de la profondeur. Cette forme de distribution exponentielle pour les racines est cohérente avec les observations faites dans les fausses pédologies à Paracou. Cependant, cela ne poserait aucune difficulté d'utiliser une autre distribution.
- Tous les paramètres inconnus dans les modèles relatifs aux caractéristiques du sol (paramètre de la distribution exponentielle des racines, points de capacité au champ et points de flétrissements pour toutes les couches) sont estimés à partir des données de pluviométrie et des sondes TDR. Nous avons formalisé le modèle dans le cadre bayésien et développé un algorithme MCMC pour l'inférer en intégrant les incertitudes de mesures de la sonde. Ce dernier point est particulièrement important car il est très complexe et très coûteux de faire des prélèvements et analyses de sol pour accéder à ces quantités. Notre modèle est ainsi assez facilement transposable à d'autres sites.

Nous avons pu facilement comparer les réponses du modèle et les profils hydriques du sol selon la topographie. Au regard des résultats d'estimation (figure



3.9), le modèle semble atteindre l'objectif d'estimer correctement l'eau disponible pour les arbres. Par ailleurs, concernant la figure 3.9, il est normal que le modèle ne puisse pas prédire les valeurs de  $REW > 1$  car cela correspond à des phénomènes d'engorgement du sol qui ne sont pas intégrés dans le modèle. Ce n'est pas du tout problématique pour nous, car nous cherchons à évaluer l'effet du stress hydrique sur la dynamique forestière. Or il est admis qu'il y a un stress lorsque le  $REW < 0.4$  et pour cette gamme de valeur le modèle prédit toujours bien.

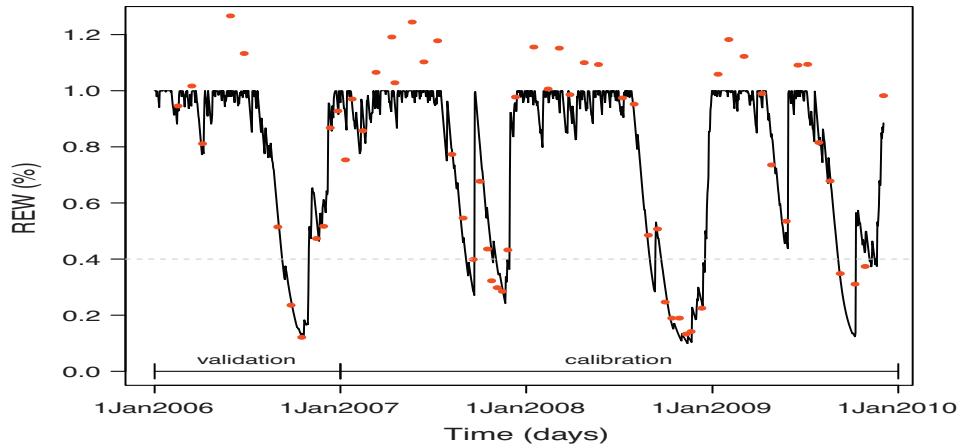


FIGURE 3.9 – REW : prédits par le modèle (—), mesurés des sondes (·) (extrait de Wagner et al. (2011))

Ayant développé un outil permettant d'évaluer directement l'eau disponible pour les arbres, nous pouvions envisager de mettre en relation cette information avec la dynamique des arbres. Cependant, il demeure une difficulté, les données des inventaires forestiers sont au mieux annuelles. La question de l'agrégation de l'indicateur est ainsi toujours présente. Pour avancer sur cette question nous avons mener une étude sur l'effet de variables climatiques, dont le REW, sur la croissance de 205 arbres appartenant à 54 espèces sur le dispositif GuyaFlux. Ces arbres étant munis de dendromètres, leurs accroissements en diamètre ont été mesurés au millimètre près en moyenne tous les 40 jours. Pour gérer les décalages entre les mesures d'accroissement et éluder la question de l'agrégation des variables climatiques mesurées quotidiennement, nous sommes partis d'un modèle de croissance journalier :

$$dbh_{i,d+1} - dbh_{i,d} = tree_i + day_d + \epsilon_{i,d} \quad \text{with } \epsilon_{i,d} \sim \mathcal{N}(0, \sigma^2) \quad (3.8)$$

où  $dbh_{i,d}$  est le diamètre de l'arbre  $i$  le jour  $d$ ,  $tree_i$  est l'effet individuel  $tree$  de l'arbre  $i$  sur la croissance journalière,  $day_d$  est l'effet du jour  $d$  sur la croissance

de tous les arbres et  $\epsilon_{i,d}$  est l'erreur du modèle supposée normale. La croissance de l'arbre  $i$  sur une période  $j$  commençant le jour  $d$  et durant  $nd_j$  jours s'obtient en sommant  $nd_j$  fois l'équation (3.8) :

$$\underbrace{dbh_{i,d+nd_j} - dbh_{i,d}}_{Gr_{i,j}} = nd_j \times tree_i + \underbrace{day_d + \dots + day_{d+nd_j-1}}_{period_j} + \underbrace{\epsilon_{i,d} + \dots + \epsilon_{d+nd_j-1}}_{\tilde{\epsilon}_{i,j}} \quad (3.9)$$

L'équation (3.9) peut s'écrire :

$$Ac_{i,j} = nd_j \times tree_i + period_j + \tilde{\epsilon}_{i,j} \quad \text{with } \tilde{\epsilon}_{i,j} \sim \mathcal{N}(0, nd_j \sigma^2) \quad (3.10)$$

où  $Ac_{i,j}$  est l'accroissement de l'arbre  $i$  sur la période  $j$ ,  $period_j$  est l'effet  $period$  pour la période  $j$ ,  $nd_j$  est le nombre de jours de la période  $j$  et  $\tilde{\epsilon}_{i,j}$  est l'erreur du modèle supposée normale. Le modèle (3.10) n'est pas une régression linéaire classique car la variance  $nd_j \sigma^2$  des termes d'erreur change selon les périodes. Il suffit de normaliser l'équation (3.10) pour obtenir des erreurs identiquement distribuées :

$$\frac{Ac_{i,j}}{\sqrt{nd_j}} = \sqrt{nd_j} \times tree_i + \frac{period_j}{\sqrt{nd_j}} + \varepsilon_{i,j} \quad \text{with } \varepsilon_{i,j} \sim \mathcal{N}(0, \sigma^2) \quad (3.11)$$

où  $\varepsilon_{i,j} = \tilde{\epsilon}_{i,j}/\sqrt{nd_j}$  et  $Var(\varepsilon_{i,j}) = Var(\tilde{\epsilon}_{i,j})/nd_j = \sigma^2$ . Ce modèle (3.11) a été notre modèle de référence. Nous avons évalué la perte en pouvoir explicatif lorsque la période est remplacée par une ou plusieurs variables climatiques(cf. table 3.2).

Nous avons tout d'abord évalué le pouvoir explicatif individuel de chaque variable climatique en ajustant les modèles  $m_{varclim}$ , avec  $varclim=REW, U^*, T min, T max, irradiance, VPD, Patm, HR$  or  $rainfall$ , définis par :

$$\frac{Ac_{i,j}}{\sqrt{nd_j}} = \sqrt{nd_j} \times tree_i + \beta_{varclim} \times \frac{varclim_j}{\sqrt{nd_j}} + \varepsilon_{i,j} \quad (3.12)$$

Ensuite nous avons cherché la combinaison de variables qui maximisaient le BIC, pour obtenir le modèle  $m_{BIC}$  :

$$\begin{aligned} \frac{Ac_{i,j}}{\sqrt{nd_j}} = & \sqrt{nd_j} \times tree_i + \beta_{REW} \frac{REW_j}{\sqrt{nd_j}} + \beta_{U^*} \frac{U_j^*}{\sqrt{nd_j}} \\ & + \beta_{T min} \frac{T min_j}{\sqrt{nd_j}} + \beta_{irradiance} \frac{irradiance_j}{\sqrt{nd_j}} + \varepsilon_{i,j} \end{aligned} \quad (3.13)$$

Il est ressorti que l'effet individuel  $tree$  expliquait 26% des variations de croissance intra-annuelle entre 205 arbres de l'étude et l'effet période expliquait seulement 9%. Parmi les variables climatiques le REW est celle qui explique le plus

Variable	Daily computation	Description	Mean	SD	Unit
<i>irradiance</i>	mean	global irradiance	590.34	170.17	$MJ.m^{-2}.d^{-1}$
<i>T min</i>	minimum	temperature minimum	23.43	0.82	celsius degree
<i>T max</i>	maximum	temperature maximum	28.37	1.35	celsius degree
<i>REW</i>	-	relative extractable water	0.77	0.28	-
<i>VPD</i>	mean	vapor pressure deficit	6.05	1.68	kPa
<i>rainfall</i>	sum	precipitation	9.03	17.21	mm
<i>U*</i>	mean	friction velocity	0.37	0.11	$m.s^{-1}$

TABLE 3.2 – Variables climatiques utilisées pour expliquer les variations de croissance intra-annuelles de 2007 à 2010, excepté pour le REW les autres variables sont mesurées toutes les demi-heures.

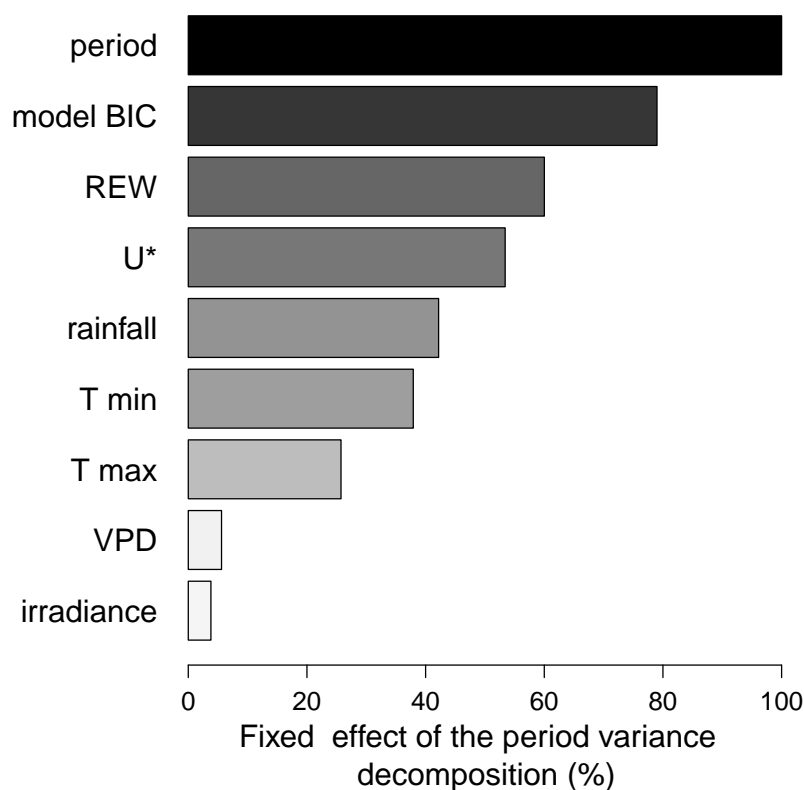


FIGURE 3.10 – Effet des variables climatiques sur les variations de croissance intra-annuelles (extrait de [Wagner et al. \(2012\)](#))

cet effet période avec 60% de la variabilité. En particulier, le REW a un pouvoir

explicatif supérieur à la simple pluviométrie, ce qui démontre l'intérêt du modèle de bilan hydrique que nous avons développé. Le modèle optimal selon le BIC, qui contient REW, explique 80% de la variabilité de l'effet période (cf. figure 3.10). Malgré le fait que l'effet période explique seulement 9% des variations de croissance, si on met en regard le REW et la croissance moyenne des arbres sur les périodes (cf. figure 3.11) le lien est clairement visible. Il me semble donc tout à fait pertinent d'utiliser prioritairement cette variable lorsque l'on veut introduire le climat dans les modèles de croissance.

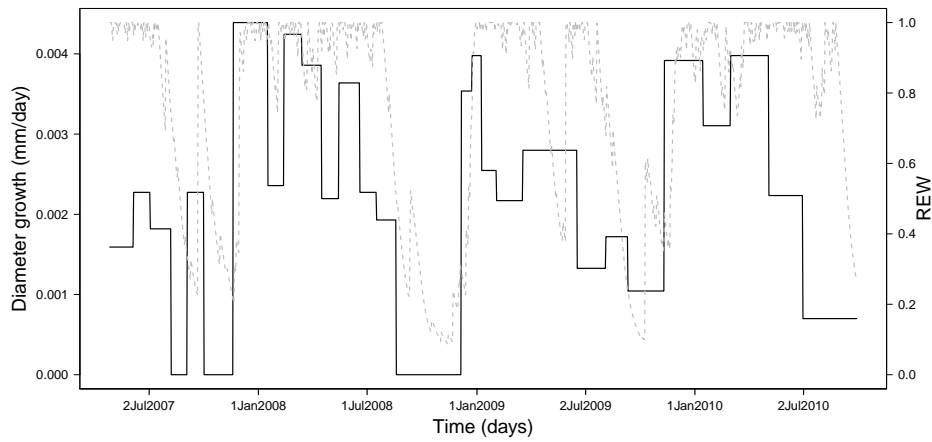


FIGURE 3.11 – Accroissements moyens des arbres sur les périodes (—); REW prédit par le modèle de bilan hydrique (· · ·) (extrait de Wagner et al. (2012))

Cette étude ayant permis d'établir un lien significatif entre les variables climatiques et la croissance des arbres, nous avons décidé de l'approfondir en nous intéressant aux effets des interactions espèces - variables climatiques sur la croissance REF. Comme dans nos précédents travaux, nous avons choisi d'utiliser les traits fonctionnels pour caractériser les espèces. Cela permettait également d'étudier les liens entre les traits et la résistance à la sécheresse. Comme pour 53 des espèces, 13 des 17 traits fonctionnels (cf. table 3.1) étaient disponibles, cela n'a pas posé de difficulté particulière en terme de modélisation. Nous avons ainsi travaillé à partir de la structure du modèle (3.11) mais en rajoutant les interactions entre traits fonctionnels et variables climatiques :

$$Ac_{i,j} = nd_j \times tree_i + period_j + nd_j \times \theta_{k,l} \times vartrait_i \times varclim_j + \epsilon_{i,j} \quad (3.14)$$

avec  $\epsilon_{i,j} \sim \mathcal{N}(0, weight \times \sigma^2)$ ,  $weight = nd_j \times n_{indSp}$  le produit entre le nombre

de jour de la période et le nombre d'individus de l'espèce de l'arbre  $i$ . Le paramètre  $\theta_{k,l}$  quantifie l'effet de l'interaction entre le  $k$ ème trait et la  $l$ ème variable climatique.

Les résultats obtenus ont été assez mitigés en terme de pouvoir explicatif de la variabilité de la croissance. L'interaction entre le 95ème percentile du DBH de l'espèce et le déficit du sol en eau (SWD=1-REW), qui avait le meilleur pouvoir explicatif, n'expliquait que 0.38% de la variance de la croissance intra-annuelle. Malgré leur faible pouvoir explicatif les effets obtenus pour ces interactions étaient globalement cohérent avec la littérature. Cette concordance suggère que ces effets sont bien réels et ne sont pas simplement contingent à l'échantillonnage. Leur faible pouvoir explicatif est peut être du au fait qu'il n'y a pas eu d'évènement climatique extrême durant la période d'observation. Les espèces présentes étant adaptées au climat local, les comportements ont du être dans les gammes de variation habituelle. Cependant, il est envisageable que les effets des interactions, bien qu'étant marginaux sur la croissance en conditions normales, caractérisent les spécificités des espèces. Selon leurs caractéristiques, les espèces devraient réagir différemment aux changements climatiques. J'envisage d'étudier, par simulations, si ces effets marginaux peuvent entraîner une évolution de la composition spécifique, selon différents scénarios climatiques.

## 4 Perspectives

Mes perspectives de travaux sont principalement orientées vers la modélisation autour des forêts tropicales. Cependant, j’ai deux pistes de recherche, relevant plus de la méthodologie, qui à mon avis méritent d’être explorées. Ces dernières sont présentées dans la première section de ce chapitre, ensuite je présente mes perspectives de recherche à court-moyen terme et à long terme.

### 4.1 Méthodologie en modélisation statistique

#### Optimisation du filtre particulière à convolution

Parmi les filtres particuliers que j’ai développés, le plus performant est le filtre particulier à convolution avec ré-échantillonnage. Ce filtre est plus souple d’utilisation et plus stable que les filtres particuliers usuels. Cependant, il est nécessaire de calibrer manuellement un paramètre,  $h$ , qui correspond à la largeur de la fenêtre du noyau de convolution utilisé pour régulariser le filtre. J’ai défini des règles pour choisir  $h$  de façon à ce que l’algorithme conserve ses propriétés de convergence. Mais en pratique, il est nécessaire de faire quelques analyses préliminaires pour trouver une valeur fonctionnelle.

La piste que j’ai commencée à explorer s’inspire de l’algorithme Exception-Maximisation (EM). A chaque itération, pour choisir  $h$  on maximise la vraisemblance des particules à l’instant suivant. Cela induit donc un décalage de un pas de temps par rapport aux observations et rallonge le temps de calcul d’une itération. Cependant, les comparaisons de performances avec les autres sur le modèle «benchmark» utilisé par [Casarin and Marin \(2009\)](#) sont très prometteuses. Avec 100 particules l’algorithme est plus performant que la plupart de filtres particuliers améliorés avec 10 000 particules, sauf pour l’Auxiliary Particle Filter (APF) ([Pitt and Shephard, 1999](#)) pour lequel il faut monter à 1000 particules.

Cette étape d'optimisation ne remet pas en cause les propriétés théoriques de convergence de l'algorithme vers le filtre optimal. Mais il est possible que cela améliore la vitesse de convergence. Une étude des propriétés théoriques de ce nouvel algorithme permettrait de le déterminer.

## Un algorithme particulier pour estimer hors-ligne les paramètres d'un modèle à espace d'état

Pour ce placer dans le cas le plus général, considérons un modèle à espace d'état (4.1) :

$$\begin{cases} x_t = f_t(x_{t-1}, \theta, \varepsilon_t) \\ y_t = h_t(x_t, \theta, \eta_t) \end{cases} \quad (4.1)$$

avec  $x_t \in \mathbb{R}^d$  les variables d'état cachées,  $y_t \in \mathbb{R}^q$  les variables observées,  $\theta \in \mathbb{R}^l$  les paramètres inconnus,  $\varepsilon_t$  les bruits de modèle et  $\eta_t$  les bruits du modèle d'observation. L'inférence bayésienne de modèles à espace d'état est souvent employée pour l'évaluation de ressources naturelles (Buckland et al., 2004; Rivot et al., 2004). Dans ces applications la série d'observations  $y_1, \dots, y_T$  est courte, de l'ordre de quelques dizaines. Il est alors possible d'employer des méthodes non séquentielles pour estimer les quantités inconnues, comme les méthodes MCMC (Gilks et al., 1996; Gimenez et al., 2007).

Le succès des méthodes MCMC est dû à leur très large domaine d'application et à leur simplicité d'implémentation. Des logiciels «boîtes noires» tels que WinBugs ont permis une très large diffusion dans les communautés non statisticiennes. Cependant, lorsque le modèle est complexe la phase où la chaîne explore l'espace avant de converger l'état stationnaire, communément appelé le temps de chauffe peut être extrêmement long. Nous avons envisagé d'utiliser le filtrage particulier, beaucoup plus rapide, pour cette phase et ainsi récupérer de bons points de départ pour l'algorithme MCMC. Mais cela n'a pas donné les résultats attendus car si le modèle est complexe, avec peu d'observation les filtres particuliers n'ont pas le temps de converger.

La piste que j'envisage d'explorer est de coupler un algorithme type filtre particulier avec ré-échantillonnage à l'optimisation de la vraisemblance. Plus concrètement, on choisit une valeur initiale pour  $\theta_0$ , on simule des trajectoires possibles pour les états cachés à l'aide d'un filtre particulier avec ré-échantillonnage puis on cherche la valeur de  $\theta$  qui maximise la vraisemblance de ces trajectoires d'états. La difficulté réside dans le caractère aléatoire de modèle (4.1) et de l'étape de ré-échantillonnage du filtre particulier. Il est possible de contourner cette difficulté

en raisonnant conditionnellement aux valeurs des variables aléatoires du modèle. Il faudrait donc répéter plusieurs fois ces étapes pour obtenir un ensemble de valeurs estimées pour  $\theta$ .

Les propriétés théoriques de cette méthode restent à étudier, mais j'ai bon espoir qu'il soit possible de démontrer que l'ensemble de valeurs estimées pour  $\theta$  est un estimateur de la loi *a posteriori* de  $\theta$ .

## **4.2 Modélisation des dynamiques forestières permettant d'évaluer l'impact des changements climatiques sur les services des forêts tropicales**

Il est à présent clairement établi ([Purves and Pacala, 2008](#)) que les simulateurs de forêts vont être les outils privilégiés pour étudier l'impact des changements climatiques sur les forêts à plus ou moins long terme.

Nous avons construit un modèle de croissance et un modèle de mortalité à l'échelle de la communauté. Nous avons réussi à introduire des variables climatiques de façon satisfaisante dans ces modèles. Les interactions entre variables climatiques et traits fonctionnels ayant des effets très faibles sur la croissance, les changements sur la composition de la communauté seront donc visibles à long terme. Les simulations me semblent être la seule alternative pour explorer l'impact de scénarios climatiques sur la forêt tropicale. Pour mettre au point un simulateur fiable de forêt tropicale dépendant du climat, il reste encore plusieurs étapes à franchir. Tout d'abord, nous n'avons pas de modèle de recrutement ou régénération à l'échelle de la communauté basé sur les traits et intégrant le climat. Nous pourrions tout à fait construire un tel modèle de recrutement en suivant la même méthodologie que pour les modèles de croissance et mortalité, mais dans un premier temps nous prévoyons d'opter pour une approche neutraliste. C'est à dire que nous allons maintenir les effectifs de la communauté constants, lorsqu'un arbre meurt il est remplacé par un arbre dont l'espèce est tiré aléatoirement parmi les espèces des arbres présents. Cette hypothèse simplificatrice, nous empêchera d'expérimenter des scénarios climatiques conduisant à la disparition de la forêt, nous pourrions uniquement explorer l'impact des scénarios climatiques sur la composition floristique. Cependant il y a plusieurs avantages à procéder de cette façon. Sur le plan pratique, il est plus simple d'avoir uniquement les deux modèles croissance et mortalité à coupler dans un simulateur car cela élimine les risques de divergence



du type explosion ou extinction du nombre d'arbres. Le risque étant alors que le simulateur peut présenter un comportement artificiellement stable sans que la dynamique forestière sous-jacente soit cohérente. Il est possible de réduire ce risque en vérifiant l'évolution de certains indicateurs de la structure d'un peuplement simulé sous le climat actuel.

Ceci dit, le couplage des modèles de mortalité et croissance basés sur les traits et intégrant le climat nécessite quelques développements méthodologiques. En effet, nous prévoyons d'utiliser l'architecture du simulateur SELVA ([Gourlet-Fleury and Houllier \(2000\)](#)) pour coupler nos modèles.

Tout d'abord, le simulateur fonctionne plutôt sur un pas de temps court, 2 ou 3 ans, alors que nous avons calibré les modèles sur des périodes longues 15-16 ans. Nous pourrions faire fonctionner SELVA sur des pas de temps de 16 ans mais cela exclurait la possibilité d'ajouter des variables de compétition, type surface terrière du voisinage, car elle perdrait toute pertinence sur de pas de temps aussi longs. Ainsi il ne serait pas possible d'étudier à terme l'effet croisé de scénarios climatiques et de l'exploitation sur la dynamique forestière. Il nous faut donc recalibrer les modèles en prenant en compte les dépendances entre les données, puisque nous avons plusieurs observations pour chaque individu. Pour le modèle de croissance cela peut se gérer par l'ajout d'un effet individuel ou d'une auto-corrélation dans l'erreur du modèle. Pour le modèle de mortalité, il faut utiliser le formalisme des modèles de survie ce qui complique un peu l'inférence.

D'autre part, nous avons inféré ces modèles séparément. Or il est possible que d'intégrer les traits dans les deux modèles engendre une sur-expression de leur effet. Pour éviter ce travers, nous envisageons d'intégrer une variable exprimant la vigueur de l'arbre dans le modèle de mortalité. C'est une pratique courante, dans notre cas nous pensons utiliser une fonction de l'écart entre la croissance théorique prédite par le modèle et la croissance observée. De cette manière, si l'utilisation des traits dans le modèle de croissance est suffisante, seule cette variable vigueur restera significative dans le modèle de mortalité. Ce couplage nécessiterait donc l'inférence simultanée des deux modèles. Cela est tout à fait envisageable dans le cadre bayésien en développant un algorithme MCMC spécifique.

Une fois que le nouveau simulateur de dynamique forestière sera opérationnel, nous générerons des données climatiques selon les différents scénarios et simulerons des évolutions du peuplement. Nous pourrions par exemple analyser comment les distributions des traits fonctionnels évoluent suivant les scénarios.

A moyen terme, il faudrait introduire également un modèle de recrutement dépen-

dant de l'état du peuplement, du climat et des traits fonctionnels. L'introduction de ce modèle dans le simulateur posera un peu plus de difficulté car il faudra coupler trois modèles et certainement inférer les trois simultanément.

Ainsi, cela permettra d'explorer l'impact des interactions entre diverses perturbations, comme l'exploitation, et les scénarios de changements climatiques. En particulier, nous pourrions explorer des scénarios climatiques extrêmes prévoyant la disparition en Guyane de la forêt tropicale humide (Cox et al., 2004).

### **4.3 Prise en compte des enjeux socio-économiques des forêts tropicales**

A plus long terme, je compte changer d'échelle de travail. Lorsque nous aurons des simulateurs de peuplements forestiers fonctionnels pour étudier l'impact de perturbations climatiques ou anthropiques sur un massif forestier, il me semble important de replacer la forêt dans son contexte socio-économique. Hormis en Guyane, où la pression sur la forêt est très faible, les forêts tropicales sont généralement au cœur d'enjeux divers (Steffan-Dewenter et al., 2007). En particulier, je suis convaincu que les forêts tropicales humides vont être plus affectées à court terme par les activités humaines que par les changements climatiques, surtout dans les pays avec une faible gouvernance.

Pour modéliser de tels socio-écosystèmes forestiers tropicaux, il est nécessaire de travailler à une échelle plus grande que celle de la parcelle expérimentale. Il faut raisonner à l'échelle d'un état ou d'une région soumis aux mêmes contraintes et aux mêmes enjeux. Ensuite, je pense qu'il faut distinguer d'une part les dynamiques spontanées et les scénarios de développement envisagés. Les dynamiques spontanées, telles que l'extension non régulée des zones urbaines ou agricoles sur la forêt, peuvent être appréhendées à l'aide de la télédétection et formalisées pour des modèles simples relativement fiables à court terme. Les prévisions faites à partir de ces modèles serviront de point de départ pour évaluer des scénarios de développements socio-économiques. Par exemple, pour la mise en place des concessions forestières on pourrait évaluer la rentabilité économique selon les variations du cours du bois et quantifier l'impact sur les services de la forêt. Pour quantifier l'impact sur la forêt, de ces différents scénarios conjugués aux dynamiques spontanées, nous pourrions utiliser les simulateurs de peuplements forestiers.

Cependant, notre simulateur de peuplement ayant été développé essentiellement à partir de données des forêts de Guyane, il y aura au préalable un travail de modéli-

sation pour l'adapter à d'autres terrains. Comme nous avons développé des méthodologies de calibration des modèles sous-jacents quasi-automatique, ce travail ne sera pas insurmontable. Mais il restera un point qui nécessitera un travail certain, l'extrapolation des estimations des services à toute la forêt du territoire étudié et l'agrégation de ces estimations nécessaire au bilan des scénarios. Pour l'extrapolation, l'utilisation de données de télédétection me semble être la voie à privilégier. Pour l'agrégation, les différences de niveau d'incertitude suivant les parties du territoire sont à intégrer rigoureusement, la méthodologie que nous avons développée pour construire la carte de biomasse de la Guyane pourra alimenter cette réflexion.

# Bibliographie

- D Alder and J. N. M Silva. An empirical cohort model for management of terra firme forests in the brazilian amazon. *Forest Ecology and Management*, 130 (1-3) :141–157, 2000. Alder, D Silva, JNM.
- M Aubry-Kientz, B Hérault, C Ayotte-Trepanier, C Baraloto, and V Rossi. Towards trait-based mortality models for tropical forests ? *PLoS ONE*, under review.
- C Baraloto, O. J Hardy, C. E. T Paine, K. G Dexter, C Cruaud, L. T Dunning, M.-A Gonzalez, J.-F Molino, D Sabatier, V Savolainen, and J Chave. Using functional traits and phylogenetic trees to examine the assembly of tropical tree communities. *Journal of Ecology*, 100(3) :690–701, 2012.
- C Baraloto, C. E. T Paine, S Patino, D Bonal, B Herault, and J Chave. Functional trait variation and sampling strategies in species-rich plant communities. *Functional Ecology*, 24(1) :208–216, 2010.
- B Barnes, D Zak, and S Denton. *Forest ecology*. Wiley, 1998.
- J. M Bernardo and A Smith. *Bayesian Theory*. Wiley, 1994.
- L Bertino, G Evensen, and H Wackernagel. Sequential data assimilation techniques in oceanography. *International Statistical Review*, 71 :223–241, 2003.
- S Bohlman and S Pacala. A forest structure model that determines crown layers and partitions growth and mortality rates for landscape-scale applications of tropical forests. *Journal of Ecology*, 100(2) :508–518, 2012.
- G. B Bonan. Forests and climate change : Forcings, feedbacks, and the climate benefits of forests. *Science*, 320(5882) :1444–1449, 2008.

- S Brooks, R King, and B Morgan. A bayesian approach to combining animal abundance and demographic data. *Animal Biodiversity and Conservation*, 27 : 515–529, 2004.
- S Buckland, K Newman, L Thomas, and N Koesters. State-space models for the dynamics of wild animal populations. *Ecological modelling*, 171 :157–175, 2004.
- F Campillo, P.Cantet, R Rakotozafy, and V Rossi. Méthodes mcmc en interaction pour l'évaluation de ressources naturelles. *revue Africaine de la Recherche en Informatique et Mathématiques Appliquées*, 8 :64–80, 2008a.
- F Campillo, R Rakotozafy, and V Rossi. Computational probability modeling and bayesian inference. *revue Africaine de la Recherche en Informatique et Mathématiques Appliquées*, 9 :123–143, 2008b.
- F Campillo, R Rkotozafy, and V Rossi. Parallel and interacting markov chain monte carlo algorithm. *Mathematics and Computers in Simulation*, 79(12) : 3424 – 3433, 2009.
- F Campillo and V Rossi. Convolution particle filter for parameter estimation in general state-space models. *IEEE Transactions on Aerospace and Electronic Systems*, 45(3) :1063–1072, 2009.
- C Canham, P LePage, and C K. A neighborhood analysis of canopy tree competition : effects of shading versus crowding. *Canadian Journal of Forest Research*, 34 :778–787, 2004.
- O Cappé, A Guillin, J. M Marin, and C. P Robert. Population monte carlo. *Journal of Computational and Graphical Statistics*, 13(4) :907–929, 2004.
- R Casarin and J Marin. Online data processing : Comparison of bayesian regularized particle filters. *Electronic journal of statistics*, 3 :239–258, 2009.
- H Caswell. *Matrix population models : construction, analysis, and interpretation*. Sinauer Associates, 2001.
- M Chaleyat-Maurel and D Michel. Des résultats de non existence de filtre de dimension finie. *Stochastics*, 13(1-2) :83–102, 1984.
- J Chase and M Leibold. *Ecological Niches : Linking Classical and Contemporary Approaches*. Interspecific Interactions. University of Chicago Press, 2003.

- J Chave, R Condit, S Aguilar, A Hernandez, S Lao, and R Perez. Error propagation and scaling for tropical forest biomass estimates. *Philosophical Transactions of the Royal Society of London B*, 359 :409–420, 2004.
- J Chave, D Coomes, S Jansen, S. L Lewis, N. G Swenson, and A. E Zanne. Towards a worldwide wood economics spectrum. *Ecology Letters*, 12(4) :351–366, 2009.
- R Condit, S. P Hubbell, and R. B Foster. Mortality-rates of 205 neotropical tree and shrub species and the impact of a severe drought. *Ecological Monographs*, 65(4) :419–439, 1995. CONDIT, R HUBBELL, SP FOSTER, RB.
- P. M Cox, R. A Betts, M Collins, P. P Harris, C Huntingford, and C. D Jones. Amazonian forest dieback under climate-carbon cycle projections for the 21st century. *Theoretical and Applied Climatology*, 78 :137–156, 2004.
- R Daley. *Atmospheric data analysis*. Cambridge University Press, 1991.
- P Del Moral. Nonlinear filtering using random particles. *Theory of Probability and its Applications*, 40(4) :690–701, 1995.
- C Delor, D Lahondère, E Egal, and P Marteau. Carte géologique de la france à 1 :500 000. département de la guyane., 2001.
- S Diaz, J Hodgson, K Thompson, M Cabido, J Cornelissen, A Jalili, G Montserrat-Martí, J Grime, F Zarrinkamar, Y Asri, S Band, S Basconcelo, P Castro-Díez, G Funes, B Hamzehee, M Khoshnevi, N Pérez-Harguindeguy, M Pérez-Rontomé, F Shirvany, F Vendramini, S Yazdani, R Abbas-Azimi, A Bogaard, S Boustani, M Charles, M Dehghan, L de Torres-Espuny, V Falczuk, J Guerrero-Campo, A Hynd, G Jones, E Kowsary, F Kazemi-Saeed, M Maestro-Martínez, A Romo-Díez, S Shaw, B Siavash, P Villar-Salvador, and M Zak. The plant traits that drive ecosystems : Evidence from three continents. *Journal of Vegetation Science*, 15(3) :295–304, 2004.
- P. J Diggle and P Ribeiro. *Model-based Geostatistics*. Springer Series in Statistics, 2007.
- A Doucet, N de Freitas, and N. J Gordon, editors. *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science. Springer-Verlag, New York, 2001.

- A Doucet and V. B Tadić. Parameter estimation in general state–space models using particle methods. *Annals of the Institute of Statistical Mathematics*, 55 (2) :409–422, 2003.
- J Ebeling and M Yasué. Generating carbon finance through avoided deforestation and its potential to create climatic, conservation and human development benefits. *Phil. Trans. R. Soc. B*, 363 :1917–1924, 2008.
- B. M. J Engelbrecht, L. S Comita, R Condit, T. A Kursar, M. T Tyree, B. L Turner, and S. P Hubbell. Drought sensitivity shapes species distribution patterns in tropical forests. *Nature*, 447 :80–82, 2007.
- V Favrichon. Classification des espèces arborées en groupes fonctionnels en vue de la réalisation d’un modèle de dynamique de peuplement en forêt guyanaise. 49 :379–403, 1994.
- A Franc, S Gourlet-Fleury, and N Picard. *Une introduction à la modélisation des forêts hétérogènes*. ENGREF, Nancy, France, 2000.
- J Gash, L C., and L G. Estimating sparse forest rainfall interception with an analytical model. *Journal of Hydrology*, 170 :79–86, 1995.
- W Gilks, S Richardson, and D Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Interdisciplinary Statistics. Chapman & Hall, 1996.
- O Gimenez, V Rossi, R Choquet, C Dehais, B Doris, H Varella, J Vila, and R Pradel. State-space modelling of data on marked individuals. *Ecological Modelling.*, 2006 :431–438, 2007.
- H Gitay and I Noble. *Plant Functional Types : Their Relevance to Ecosystem Properties and Global Change*, chapter What are functional type and how should we seek them ?, pages 3–18. Cambridge University Press, 1997.
- S Goetz, A Baccini, N Laporte, T Johns, W Walker, J Kellndorfer, R Houghton, and M Sun. Mapping and monitoring carbon stocks with satellite observations : a comparison of methods. *Carbon Balance and Management*, 4(1) :2, 2009.
- V Gond, V Freycon, J.-F Molino, O Brunaux, F Ingrassia, P Joubert, J. F Pekel, M.-F Prévost, V Thierron, P. J Trombe, and D Sabatier. Broad-scale spatial pattern of forest landscape types in the Guiana Shield. *International Journal of Applied Earth Observation and Geoinformation*, 13 :357–367, 2011.

- S Gourlet-Fleury. *Modélisation individuelle spatialement explicite de la dynamique d'un peuplement de forêt dense tropicale humide (dispositif de Paracou - Guyane française)*. Thèse de doctorat, Université Claude Bernard - Lyon I, Lyon, France, 1997. 2 volumes.
- S Gourlet-Fleury, G Cornu, S Jesel, H Dessard, J. G Jourget, L Blanc, and N Picard. Using models to predict recovery and assess tree species vulnerability in logged tropical forests : A case study from french guiana. *Forest Ecology and Management*, 209(1-2) :69–86, 2005. Gourlet-Fleury, S Comu, G Jesel, S Dessard, H Jourget, JG Blanc, L Picard, N Conference on Meeting the Challenge - Silvicultural Research in a Changing World JUN 14-18, 2004 Montpellier, FRANCE Int Union Forest Res Org, Div 1, CIRAD Foret, INRA, USDA Forest Serv.
- S Gourlet-Fleury and F Houllier. Modelling diameter increment in a lowland evergreen rain forest in french guiana. *Forest Ecology and Management*, 131 : 269 – 289, 2000.
- S Gourlet-Fleury, V Rossi, M Rejou-Mechain, V Freycon, A Fayolle, L Saint-André, G Cornu, J Gérard, J Sarrailh, O Flores, F Baya, A Billand, N Fauvet, M Gally, M Henry, D Hubert, A Pasquier, and N Picard. Environmental filtering of dense-wooded species controls above-ground biomass stored in african moist forests. *Journal of Ecology*, 99(4) :981–990, 2011.
- A Granier, N Bréda, P Biron, and S Villette. A lumped water balance model to evaluate duration and intensity of drought constraints in forest stands. *Ecological Modelling*, 116(2-3) :269 – 283, 1999.
- B Hérault. Reconciling niche and neutrality through the emergent group approach. *Perspectives in Plant Ecology, Evolution and Systematics*, 9(2) :71 – 78, 2007.
- B Hérault, B Bachelot, L Poorter, V Rossi, F Bongers, J Chave, C. E Paine, F Wagner, and C Baraloto. Functional traits shape ontogenetic growth trajectories of rain forest tree species. *Journal of Ecology*, 99 :1431–1440, 2011.
- S. P Hubbell. *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton University Press, 2001.
- IPCC. *Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Geneva, Switzerland, core writing team, pachauri, r.k. and reisinger edition, 2007.



- R. E Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME—Journal of Basic Engineering*, 82, Series D :35–45, 1960.
- L Kuo and B Mallick. Variable selection for regression models. *Sankhya Ser.B*, (60) :65–81, 1998.
- S Lavorel and E Garnier. Predicting changes in community composition and ecosystem functioning from plant traits : revisiting the holy grail. *Functional Ecology*, 16 :545–556, 2002.
- S Lavorel, S McIntyre, J Landsberg, and T Forbes. Plant functional classifications : from general groups to specific groups based on response to disturbance. *Trends in Ecology & Evolution*, 12(12) :474 – 478, 1997.
- J Lescure, H Puig, D. L B. Riera, A Beekman, and A Beneteau. La phytomasse épigée d’une forêt dense en guyane française. *Acta Oecologia*, 4, 1983.
- S Lewis, O Phillips, T Baker, J Lloyd, Y Malhi, S Almeida, N Higuchi, W Laurance, D Neill, N Silva, J Terborgh, A. T Lezama, R. V M, S Brown, J Chave, C Kuebler, P Núñez, B. V. . . I press. 359, and 421-436. Concerted changes in tropical forest structure and dynamics : Evidence from 50 south american long-term plots. *Philosophical Transactions of the Royal Society of London B*, 359 :421–436, 2004.
- Y Malhi, T Baker, O Phillips, S Almeida, E Alvarez, L Arroyo, J Chave, C Czimeczik, A. D Fiore, N Higuchi, T Killeen, S Laurance, W Laurance, S Lewis, L Montoya, A Monteagudo, D Neill, P. N V., S Patiño, N Pitman, C Quesada, J Silva, A Lezama, R. V Martinez, J Terborgh, B Vinceti, , and J Lloyd. The above-ground coarse wood productivity of 104 neotropical forest plots. *Global Change Biology*, 10 :563–591, 2004.
- E Marcon, B Hérault, C Baraloto, and G Lang. The decomposition of shannon’s entropy and a confidence interval for beta diversity. *Oikos*, 121(4) :516–522, 2012.
- J.-M Marin and C. P Robert. *Bayesian Core : A Practical Approach to Computational Bayesian Statistics*. Springer, 2007.
- A Martin and S Thomas. A reassessment of carbon content in tropical trees. *PLoS ONE*, 6(8) :e23533, 2011.

- P McCullagh and J Nelder. *Generalized Linear Models*. Chapman & Hall, 1989.
- G McLachlan and T Krishnan. *The EM Algorithm and Extensions*. Wiley Series in Probability and Statistics. Wiley, 2007.
- K Mengersen and C Robert. *Bayesian Statistics 7*, chapter The pinball sampler. Oxford University Press, 2003.
- Q Molto, V Rossi, and L Blanc. Error propagation in biomass estimation in tropical forests. *Methods in Ecology and Evolution*, in press, 2013.
- F Mortier, V Rossi, G Guillot, S Gourlet-Fleury, and N Picard. Population dynamics of species-rich ecosystems : the mixture of matrix population models approach. *Methods in Ecology and Evolution*, in press, 2013.
- N Myers, R. A Mittermeier, C. G Mittermeier, G. A. B da Fonseca<sup>3</sup>, and J Kent. Biodiversity hotspots for conservation priorities. *Nature*, 403 :853–858, 2000.
- D. C Nepstad, P Moutinho, M. B Dias-Filho, E Davidson, G Cardinot, D Markewitz, R Figueiredo, N Vianna, J Chambers, D Ray, J. B Guerreiros, P Lefebvre, L Sternberg, M Moreira, L Barros, F. Y Ishida, I Tohlver, E Belk, K Kalif, and K Schwalbe. The effects of partial throughfall exclusion on canopy processes, aboveground production, and biogeochemistry of an amazon forest. *Journal of Geophysical Research : Atmospheres*, 107(D20) :LBA 53–1–LBA 53–18, 2002.
- J Pastor and W Post. Response of northern forests to  $CO_2$ -induced climate change. *Nature*, 334 :55–58, 1988.
- O Phillips, T Baker, L Arroyo, N Higuchi, T Killeen, W. F Laurance, S. L Lewis, J Lloyd, Y Malhi, A Monteagudo, D Neill, P. N Vargas, N Silva, J Terborgh, R. V Martínez, M Alexiades, S Almeida, S Brown, J Chave, J. A Comiskey, C. I Czimczik, A. D Fiore, T Erwin, C Kuebler, S. G Laurance, H. E. M Nascimento, W Palacios, S Patiño, N Pitman, J Olivier, C. A Quesada, M Saldias, A. T Lezama, and B Vinceti. Pattern and process in amazon tree turnover, 1976–2001. *Philosophical Transactions of the Royal Society of London B*, 359 :381–408, 2004a.
- O. L Phillips, P Hall, A. H Gentry, S. A Sawyer, and R VÃ¡squez. Dynamics and species richness of tropical rain forests. *Proceedings of the National Academy of Sciences*, 91(7) :2805–2809, 1994.

- P. D Phillips, C. P de Azevedo, B Degen, I. S Thompson, J. N. M Silva, and P. R van Gardingen. An individual-based spatially explicit simulation model for strategic forest management planning in the eastern amazon. *Ecological Modelling*, 173(4) :335–354, 2004b. Phillips, PD de Azevedo, CP Degen, B Thompson, IS Silva, JNM van Gardingen, PR.
- N Picard and A Franc. Aggregation of an individual-based space-dependent model of forest dynamics into distribution-based and space-independent models. *Ecological Modelling*, 145(1) :69–84, 2001.
- N Picard, P Köhler, F Mortier, and S Gourlet-Fleury. A comparison of five classifications of species into functional groups in tropical forests of french guiana. *Ecological Complexity*, 2012.
- N Picard, F Mortier, V Rossi, and S Gourlet-Fleury. Clustering species using a model of population dynamics and aggregation theory. *Ecological Modeling*, 221 :152–160, 2010.
- M. K Pitt and N Shephard. Filtering via simulations : auxiliary particle filter. *Journal of the American Statistical Association*, 94(446) :590–599, 1999.
- D Purves and S Pacala. Predictive models of forest dynamics. *Science*, 320(5882) : 1452–1453, 2008.
- A Raftery, M Newton, J Satagopan, and P Krivitsky. *Bayesian Statistics 8*, chapter Estimating the Integrated Likelihood via Posterior Simulation Using the Harmonic Mean Identity (with Discussion), pages 1–45. Oxford University Press, 2007.
- S Richardson and P. J Green. On bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 59(4) :731–792, 1997.
- E Rivot, E Prévost, E Parent, and J Baglini-Ère. A bayesian state-space modeling framework for fitting a salmon stage-structured population dynamic model to multiple time series of field data. *Ecological Modelling*, 179(4) :463 – 485, 2004.
- C. P Robert. *Méthodes de Monte Carlo par Chaînes de Markov*. Statistique Mathématique et Probabilité. Economica, Paris, 1996.

- V Rossi and J.-P Vila. Approche non paramétrique du filtrage de système non linéaire à temps discret et à paramètres inconnus. *Comptes Rendus de l'Académie des Sciences, Série I, Mathématique*, 340(10) :759–764, 2005a.
- V Rossi and J.-P Vila. Filtrage de bioprocédé de dépollution. Approche par convolution particulière. *e-STA*, 2(1), 2005b.
- V Rossi and J.-P Vila. Bayesian multioutput feedforward neural network selection : a conjugate prior approach. *IEEE Transactions on Neural Networks*, 17(1) :35–47, 2006a.
- V Rossi and J.-P Vila. Nonlinear filtering in discrete time : A particle convolution approach. *Ann. I.S.U.P.*, 50(3), 2006b.
- V Rossi and J.-P Vila. Bayesian selection of multiresponse nonlinear regression model. *Statistics*, 42(4) :291–311, 2008.
- I Steffan-Dewenter, M Kessler, J Barkmann, M. M Bos, D Buchori, S Erasmi, H Faust, G Gerold, K Glenk, S. R Gradstein, E Guhardja, M Harteveld, D Hertel, P Höhn, M Kappas, S Köhler, C Leuschner, M Maertens, R Marggraf, S Migge-Kleian, J Mogeia, R Pitopang, M Schaefer, S Schwarze, S. G Sporn, A Steingrebe, S. S Tjitrosoedirdjo, S Tjitrosoemito, A Twele, R Weber, L Woltmann, M Zeller, and T Tschardtke. Tradeoffs between income, biodiversity, and ecosystem functioning during tropical rainforest conversion and agroforestry intensification. *Proceedings of the National Academy of Sciences*, 104(12) : 4973–4978, 2007.
- M Stone. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B*, 36(2) :111–147, 1974.
- M Swaine and T Whitmore. On the definition of ecological species groups in tropical rain forests. *Vegetatio*, 75 :81–86, 1988.
- M Uriarte, R Condit, C. D Canham, and S. P Hubbell. A spatially explicit model of sapling growth in a tropical forest : does the identity of neighbours matter ? *Journal of Ecology*, 92(2) :348–360, 2004.
- M. B Usher. A matrix model for forest management. *Biometrics*, 25 :309–315, 1969.

- P Van Leeuwen. A variance-minimizing filter for large-scale applications. *American Meteorological Society*, 131 :2071–2084, 2003.
- F Wagner, B H  rault, C Stahl, D Bonal, and V Rossi. Modeling water availability for trees in tropical forests. *Agricultural and Forest Meteorology*, 151(9) :1202–1213, 2011.
- F Wagner, V Rossi, C Stahl, D Bonal, and B H  rault. Water availability is the main climate driver of neotropical tree growth. *PLoS ONE*, 7(4) :e34074, 2012.
- M West. Approximating posterior distributions by mixtures. *Journal of the Royal Statistical Society, Series B*, 55(2) :409–422, 1993.
- B Zeide. Analysis of growth equations. *Forest Science*, 39 :594–616, 1993.